A nanoscale experiment measuring gravity’s role in breaking the unitarity of quantum dynamics

BY JASPER VAN WEZEL1,* AND TJERK H. OOSTERKAMP2

1 Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge, UK
2 Leiden Institute of Physics, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

Modern, state-of-the-art nanomechanical devices are capable of creating spatial superpositions that are massive enough to begin to experimentally access the quantum to classical crossover, and thus force us to consider the possible ways in which the usual quantum dynamics may be affected. One recent theoretical proposal describes the crossover from unitary quantum mechanics to classical dynamics as a form of spontaneous symmetry breaking. Here, we propose a specific experimental set-up capable of identifying the source of unitarity breaking in such a mechanism. The experiment is aimed specifically at clarifying the role played by gravity, and distinguishes the resulting dynamics from that suggested by alternative scenarios for the quantum to classical crossover. We give both a theoretical description of the expected dynamics, and a discussion of the involved experimental parameter values and the proposed experimental protocol.

Keywords: micromechanical devices; magnetic devices; quantum foundations; gravity

1. Introduction

Experimentalists are pushing to cool cantilevers with low intrinsic damping to ever lower temperatures, both in the context of magnetic resonance force microscopy (MRFM) (Degen et al. 2009), and in the quest to cool a mechanical resonator to its zero point motion (Cho 2003; LaHaye et al. 2004; O’Connell et al. 2010) and eventually bring it into a quantum mechanical superposition (Marshall et al. 2003; LaHaye et al. 2004, 2009; Naik et al. 2006). One such experiment has recently succeeded in cooling a resonator to its quantum mechanical ground state and has demonstrated the controllable creation of single quantum excitations (O’Connell et al. 2010). Several other feasible experiments have been proposed in which a mechanical resonator is coupled with a quantum system in such a way that, with some experimental progress, a superposition of spatially separated states of the resonator may be detectable (Marshall et al. 2003; Naik et al. 2006). With the advent of such experiments, involving both quantum mechanics and

*Author’s current address and address for correspondence: Materials Science Division, Argonne National Laboratory, Argonne, IL 60439, USA (jvanwezel@anl.gov).
mesoscopic objects, the problem of reconciling quantum mechanics with classical physics is no longer a purely theoretical endeavour, but rather becomes an experimental necessity.

Many theoretical proposals exist for how to explain the apparent absence of (spatial) quantum superpositions of macroscopic objects in the classical world. The most well known of these include the idea that environment-induced decoherence hides the superposed states from our view (Caldeira & Leggett 1983; Joos & Zeh 1985; Prokof’ev & Stamp 2000; Adler 2003; Zurek 2003); the idea that our personal participation in superpositions implies an inability to observe alternative branches of the superposed state of the universe (Everett 1957; Deutsch 1999); and the idea that corrections to Schrödinger’s equation become important at a scale intermediate between that of microscopic particles and macroscopic objects (Pearle 1976, 2009; Ghirardi et al. 1986; Pearle & Squires 1994; Bassi & Ghirardi 2003). Recently, it has been pointed out by several authors that gravity may play an important role in scenarios of the latter kind (Diósi 1987, 1989; Penrose 1996, 1998; Bassi & Ghirardi 2003; van Wezel et al. 2008).

One particular very recent proposal posits that the transition between quantum dynamics at the microscopic scale and classical physics at the macroscopic scale may be described as a form of spontaneous symmetry breaking, in direct analogy to the breaking of translational symmetry in macroscopic crystals, rotational symmetry in macroscopic magnets, and so on (van Wezel 2010). The particular symmetry being broken in this case is the unitarity of quantum mechanical time evolution, and the necessary symmetry breaking field may be supplied by the subtle influence of gravity at mesoscopic length scales.

Here, we present a specific experiment which we believe is capable of producing a superposition that is massive enough to test the theory of spontaneously broken unitarity, while allowing enough experimental control to also differentiate its predictions from those of alternative scenarios. The suggested set-up is based on realistic estimates of experimental parameters, which may be obtained in state-of-the-art MRFM experiments. The classical object being forced into a superposition is the micromechanical resonator, which forms the detector arm of a typical MRFM experiment. This set-up differs from similar proposals in the literature (Marshall et al. 2003; Naik et al. 2006) because of the entanglement of the resonator with a nearby microscopic spin state, which allows a definitive differentiation between the effects of decoherence, spontaneously broken unitarity and alternative scenarios, using the specific experimental protocol described here.

In the following, we first describe the proposed experimental set-up, and discuss the parameters that will have to be fulfilled in order for the experiment to be successful. We then present numerical simulations of the expected time evolution of this set-up in the context of spontaneously broken unitarity, which explicitly show how gravity may influence the quantum dynamics of the resonator. We also show how a pointer basis (which defines the possible outcomes of a measurement on a quantum system) and Born’s rule (which deals with the probability that such a measurement yields a given result) are automatically recovered from the time evolution of the resonator in this scenario. Finally, we turn to a detailed description of the experimental protocol required to distinguish the specific time evolution discussed here from both more common disturbances to quantum dynamics, such as decoherence, and from alternative theoretical scenarios.

Proc. R. Soc. A (2012)
A nanoscale experiment

2. The experimental set-up

Inspired by the experiment by Rugar et al. (2004), in which the force exerted by an electron spin on a small magnet is detectable by measuring the deflection of a mechanical resonator, we have proposed that such an experimental configuration can be used to bring a significant mass in a superposition involving a large displacement (van Wezel et al. 2008). The adaptation of such a set-up shown in figure 1 consists of a thin wire holding a plate of mass \( m \) as well as a small spherical magnet with magnetization \( M \), in close vicinity to a single, isolated electron spin. A natural candidate for the electron spin is the well-known Nitrogen-Vacancy (NV) colour centre in diamond because already at room temperature it has shown a coherence time as large as 250 \( \mu \)s (Takahashi et al. 2008). In isotopically engineered diamond, this increases to 1.8 \( \mu \)s (Balasubramanian et al. 2009), while at cryogenic temperatures, the \( T_2 \) time may be expected to become even larger. Rabl et al. (2009) have developed a purely quantum mechanical description of this mechanical resonator coupled to a single electron spin. For independent detection of both the cantilever and the electron spin, we envisage that the cantilever motion can be observed by coupling it to a SQUID through a coil, while the electron spin can be characterized optically. When no current is injected into the SQUID and when no light is coupled to the NV centre, the purely quantum mechanical description of Rabl et al. should normally be applicable.

While the non-magnetic \( |S_z = 0\rangle \) ground state of the spin leaves the resonator untouched, the \( |S_z = -1\rangle \) (and \( |S_z = +1\rangle \)) excited state will attract (and repel) the cantilever with a force \( F = \mu G \), where \( G \) is the magnetic field gradient originating from the magnetic sphere at the position of the electron spin with magnetization \( \mu \). Inducing transitions between and superpositions of the different spin states can readily be achieved by applying an appropriate combination of static \( B_0 \) and radio frequency \( B_1 \) magnetic fields. Starting from a superposed spin state, the magnetic coupling between the spin state and the motion of the resonator then eventually yields a superposition of two out of phase oscillation modes of the resonator.

Figure 1. The proposed experimental configuration. A thin gold plate is attached to a nanowire or multi-walled carbon nanotube to create a massive mechanical resonator. A small spherical magnet acts as a detector for the spin magnetic moment in a diamond NV centre. The graph on the right shows the energy levels of the spin as a function of magnetic field. (Online version in colour.)
For the experimental protocol described in detail in §4, and employing the realistic parameters discussed presently, the amplitude of the two anti-phase oscillations could be caused to approach the thickness of the plate. Importantly, these amplitudes could, in principle, be achieved in a time shorter than the expected dephasing time of the electron spin owing to nuclear spins in the diamond lattice (Balasubramanian et al. 2009) and shorter than the dephasing time of the cantilever owing to coupling with the phonon bath (Müller-Ebhardt et al. 2008, 2009). Thus, a superposition of the plate over macroscopically distinct positions may be achieved within the limits set by decoherence.

(a) Order of magnitude estimates for experimental parameters

The experimental parameters required to be able to unambiguously assign an observed decay time to any process other than decoherence lie only just beyond what is already being used in present-day MRFM measurements. In the experiments of Rugar et al., the deflection of a resonator owing to the interaction with a single electron spin was measured (Rugar et al. 2004). The field gradient in these experiments was $\frac{\partial B}{\partial x} = 2 \times 10^5 \text{T m}^{-1}$ which, with a magnetic moment of the electron spin of $\mu_B = 9.3 \times 10^{-24} \text{JT}^{-1}$, leads to a force $F_{\text{spin}} = \mu_B \frac{\partial B}{\partial x} = 1.8 \times 10^{-18} \text{N}$. With the stiffness of the resonator $k = 1.1 \times 10^{-4} \text{N m}^{-1}$, this would imply a static deflection of only $d_0 = F_{\text{spin}} / k = 1.7 \times 10^{-14} \text{m}$.

To increase the deflection of the resonator, the electron spin was inverted twice during each resonator period, $T_{\text{res}}$. If the spin inversions remain coherent with the resonator motion long enough, the amplitude could, in principle, be increased to $d = QF_{\text{spin}} / k$, where $Q$ is the Q-factor of the resonator. Since, in practice, the spin inversions remain coherent with the cantilever motion only for a time $\tau_m \ll QT_{\text{res}}$, the maximum amplitude is limited to $d = (1 - \exp(-\tau_m / QT_{\text{res}})) \times QF_{\text{spin}} / k$, where $\tau_m$ is the so-called rotating frame relaxation time. For the experiment by Rugar et al., $\tau_m = 760 \text{ms}$ and the maximum amplitude would exceed 60 pm (Degen et al. 2009).

For a spin in a quantum superposition of states, $\tau_m$ would have to be replaced by a dephasing time $T_2$, which has been measured to be 1.2 ms in NV centres in diamond in which the $^{13}$C isotopic content was reduced (Balasubramanian et al. 2009). Although this is the longest decoherence time measured in a solid-state system to date, the fact that $T_2$ is so much shorter than $\tau_m$ still limits the distance between the two centres of mass involved in the superposition. Fortunately, in the context of nuclear MRFM, the group of Rugar now employs field gradients of $4 \times 10^6 \text{T m}^{-1}$ (Degen et al. 2009), which when applied to an electron spin experiment, would lead to a 20-fold increase of the deflections. An alternative way of enhancing the deflection would be to employ softer springs. In our laboratory (T. H. Oosterkamp 2011, personal communications), we have been able to fabricate nanowire resonators with spring constants down to $k = 1 \times 10^{-6} \text{N m}^{-1}$. Such a nanowire, ending in a thin gold foil measuring $5 \mu\text{m} \times 5 \mu\text{m} \times 20 \text{nm}$, would have a resonance frequency of 1.6 kHz.

Combining an increase of the NV-centre dephasing time to $T_2 = 50 \text{ms}$ (which might be achievable by further eliminating nuclear spins in the diamond or by working at sufficiently low temperatures to freeze out the nuclear spins) with the highest field gradients in the Rugar laboratory and a soft $k = 1 \times 10^{-6} \text{N m}^{-1}$ resonator would result in a deflection of $d = \left( \frac{T_2}{T_{\text{res}}} \right) \times (\mu_B / k) \times \frac{\partial B}{\partial x} = 3 \text{nm}$.  

Proc. R. Soc. A (2012)
As explained in §3, the typical energy scale involved in the process of spontaneous unitarity breaking may be expected to be $\Delta = Gm^2d^2/L^3 = 4 \times 10^{-34}$ J (with $L$ the sample thickness and $G$ the gravitational constant), which yields a typical time scale $\hbar/\Delta = 25$ ms. Since this is of the same order of magnitude as the dephasing time of the NV-centre, it may be possible for the loss of unitarity to become apparent before the NV-centre decoheres.

The dephasing of the cantilever itself, on the other hand, may be an issue. This time scale can be estimated by describing the environment as an infinite bath of harmonic oscillators and integrating out the environmental degrees of freedom (Zurek 1991; Kleckner et al. 2008). For the specific case in which the dephasing is due to the phonons that are excited within the resonator, this problem has received some attention in the context of gravitational wave detectors and the time scale resulting after the phonon bath has been integrated out was shown to be $\tau_{\text{res}} = QT_{\text{res}}/(2\pi N_{\text{phon}})$, where $N_{\text{phon}} = k_B T_{\text{mode}}/\hbar f_{\text{res}}$ is the number of phonons of the resonator mode, which is determined solely by the resonator’s thermodynamic mode temperature $T_{\text{mode}}$, and frequency $f_{\text{res}}$. In order for this dephasing time to exceed $\hbar/\Delta$, the temperature of a resonator with resonance frequency $f_{\text{res}}$ and $Q=10^5$ would have to be smaller than 100 $\mu$K, which would require the experiment to be done in a nuclear demagnetization refrigerator. However, although before we considered the damping $\gamma = k/(2\pi f Q) = 1 \times 10^{-15}$ Ns m$^{-1}$ with $Q=10^5$, it is possible that future resonators will have even higher $Q$ since much lower damping values have been achieved in the carbon nanotube resonators fabricated by Huttel et al. (2009), who estimate $\gamma = 2\pi f_{\text{res}} m/Q = 2 \times 10^{-17}$. If such a low damping can be achieved for the resonator proposed here, the mode temperature may be around 1 mK without dominating the dephasing time of the NV-centre.

It has been argued before that the thermal contribution to the dephasing time considered above should give a correct order of magnitude estimate even in the presence of a driving force (Marshall et al. 2003). However, if we consider a dephasing mechanism whose strength depends explicitly on the amplitude of the oscillation as well as its mode temperature (Unruh & Zurek 1989), we should replace $N_{\text{phon}}$ in the expression for $\tau_{\text{res}}$ by $(1/2)kd^2/\hbar f_{\text{res}}$. With the lowest damping considered ($\gamma = 2 \times 10^{-17}$) at a temperature of 100 $\mu$K, this results in a dephasing time of the order of 1 ms. In that case, one would need to resort to using an even thinner gold plate to see the effect of the spontaneously broken unitarity. A gold mass of 10 nm thickness (f.e. deposited on a single sheet of graphene), in the presence of such low damping, gives rise to the time scale $\hbar/\Delta = 20$ ms, which might just be detectable within the time limit set by decoherence.

Although the phonon contribution considered here is expected to dominate the dephasing of the resonator, other mechanisms are also present. Defects in the cantilever, for example, may act as effective two-level systems at low temperatures, and the clamping of the cantilever may lead to additional phonon radiation, both contributing to the dephasing of the resonator (Remus et al. 2009; Vinante et al. 2011). The former issue can be addressed by the application of small additional magnetic fields or by reducing the number of defects in the carbon nanotube resonator, while a scheme to circumvent the problem of clamping loss using optical trapping of a resonator has recently been proposed in a different context (Li et al. 2011).
3. Simulating the time evolution

In this section, we will focus on spontaneously broken unitarity as one particular scenario for the transition from quantum dynamics to classical physics, and show explicitly how its predictions may be expected to influence the dynamics of the coupled resonator–spin system. How the proposed experimental set-up may be used to also distinguish these particular predictions from those of other scenarios will be discussed in the next section. A detailed discussion of the theory of spontaneously broken unitarity can be found in van Wezel (2010), and we will only repeat its main results here.

The basic observation underlying the idea of spontaneous unitarity breaking is the fact that the dynamics generated by the usual time-dependent Schrödinger equation is unstable, in the sense that even an infinitesimally weak perturbation may qualitatively change its behaviour in the thermodynamic limit. This situation is analogous to that of other spontaneously broken symmetries: the Hamiltonian for a crystal is invariant under translations, but the sensitivity to even infinitesimal perturbations allows a macroscopic crystal to nonetheless localize in only a single position (Anderson 1972). In the same way, the unitarity of quantum mechanical time evolution is sensitive to even infinitesimally small non-unitary perturbations, which allows sufficiently large objects to undergo classical dynamics (van Wezel 2010). An important caveat in this argument is that there has to exist a fundamental non-unitary interaction somewhere in nature (van Wezel 2008). It has been pointed by several authors that gravity may fill this role, and that in fact the energy scale on which the effects of gravity as a non-unitary influence become important, lies in the regime in which we expect the quantum to classical crossover to take place (Diósi 1987, 1989; Penrose 1996, 1998; van Wezel et al. 2008).

Because of the relatively small mass and low density of the resonator in the proposed experiment, its dynamics will be very close to purely quantum mechanical, and we expect to be able to include any non-unitary effects owing to gravity as minor perturbations to Schrödinger’s equation. Following the procedure outlined in the appendix, we thus write the dynamics of the resonator in the presence of a non-unitary gravitational perturbation as:

$$\frac{d}{dt} \psi = -\frac{i}{\hbar} \left( \hat{H} - i G \frac{m^2}{2L^3} [\hat{x} - \xi]^2 \right) \psi,$$

where $\hat{H}$ is the usual quantum mechanical Hamiltonian, $G$ is the gravitational constant and $m$ and $L$ are the mass and width of the massive plate. The operator $\hat{x}$ is the position operator for the centre of mass of the plate (measured with respect to the centre of mass of the initial wave function). The time-dependent, randomly fluctuating variable $\xi$ has been introduced as a correction to $\hat{x}$, because the theory of gravity (i.e. general relativity) insists that any quantity that can be used as a measure of distance between locations in different components of a space–time superposition must be ill-defined.

Notice that both general relativity and quantum mechanics are in fact fully deterministic theories. There is thus no reason to believe that any part of the interplay between quantum mechanics and gravity should be anything but deterministic. The introduction of a random variable in equation (3.1) should be
seen only as a poor man’s approach towards simulating an essentially ill-defined quantity: close to the point where unitary quantum mechanics is still a good description of nature, we may get away with using the concept of superpositions—even though they really are ill-defined notions in general relativity—if we include also an effectively random correction to the differences in distance between superposed spacetimes. Although superpositions and random variables may not actually feature in the exact reality of quantum gravity, we assume that if we insist on the possibility of effectively describing the state of the system as a superposition in some limit, then we also need to take into account an effectively random correction to the notion of position. We are thus led to the effective, phenomenological description of the first-order correction to the unitary Schrödinger equation given by equation (3.1). In the appendix, we give a more detailed discussion of both the steps leading to equation (3.1) and the implications of the imaginary term on the conservation of energy and on the conservation of the norm of the wave function.

In contrast to the Schrödinger–Newton equation, which has been formulated in a similar context by other authors (Diósi 1984; Moroz et al. 1998; van Wezel & van den Brink 2008), equation (3.1) is a purely linear equation, which arises naturally from the extension of the well-studied equilibrium mechanism of spontaneous symmetry breaking to the dynamical realm, and obeys all the requirements of a theory of spontaneous symmetry breaking: for unitarity to be spontaneously broken, one necessarily needs to invoke a singular thermodynamic limit, a macroscopic order parameter and a symmetry breaking field (van Wezel 2010). It is because of this connection to spontaneous symmetry breaking, that the time scale \( Gm^2 \langle x^2 \rangle / 2L^3 \), which has been suggested before to set the appropriate time scale for the influence of gravity on quantum mechanical time evolution (Diósi 1987, 1989; Penrose 1996, 1998; van Wezel et al. 2008), may be embedded into the explicit expression of the wave function dynamics provided by equation (3.1). This final form of the modified Schrödinger equation can be straightforwardly integrated using standard numerical methods to yield a phenomenological prediction for the expected time evolution of our resonator experiment.

(a) Two time scales

To be explicit, we first express the state of the system depicted in figure 1 by the quantum numbers for the position of the cantilever’s centre of mass \( x \) and the orientation of the spin’s magnetic moment \( \sigma \). We are then interested in the dynamics of the system starting from the initial state \( \psi(x, t=0) = \alpha \chi_1 \psi^0(x + d) + \beta \chi_1 \psi^0(x - d) \), where \( \psi^0(x) \) is the ground state wave function of the harmonic oscillator of mass \( m \) centred at \( x = 0 \), and \( \chi_1 \) indicates the spin state with \( S^z = \sigma \). The entangled initial state is prepared by performing an MRFM measurement on a suitably initialized superposition of the spin state, as discussed in §4.

For the purpose of clarity, we ignore the imposed oscillatory motion of the cantilever in this simulation and instead assume that it is fixed at its maximum displacement \( \pm d \). It is straightforward to also include these oscillations, but they shroud the role of the unusual dynamics induced by the phenomenological gravitational term and do not change our conclusions. Further ignoring, for the moment, the phonons and other internal degrees of freedom of the oscillator,
Figure 2. The two reduction processes induced by the modified Schrödinger equation. Starting at $t = 0$ with the superposed wave function $\sqrt{1/2}(\chi_\uparrow \psi^0(x+d) + \chi_\downarrow \psi^0(x-d))$ and $\xi > 0$, the time evolution induced by equation (3.1) with a static $\xi$ initially causes the fast reduction of the wave function to just the component $\chi_\downarrow \psi^0(x-d)$ (solid lines, $t = 0 \ldots 1\,\text{ms}$), and then the slow reduction of that component towards a wave function centred at $x = \xi$ (dashed lines, $t = 1 \ldots 20\,\text{ms}$). For clarity, all curves have been renormalized before plotting them. The parameters used in this simulation are based on the optimistic order of magnitude estimates discussed in §2. They are $m = 1 \times 10^{-14}\,\text{kg}$, $L = 1\,\text{nm}$ and $d = 1\,\text{nm}$, resulting in an initial reduction time of the order of $t = 1\,\text{ms}$. (Online version in colour.)

we are left with only the kinetic energy of the oscillator as a whole to set the Hamiltonian in equation (3.1), so that we have $\hat{H} = \hat{p}^2/(2m)$. The spin in the diamond NV centre is assumed to be free for the duration of the experiment.

The time-dependent wave function of the resonator found by solving the differential equation (3.1) now shows two distinct processes that happen simultaneously. These can be most easily understood for the special case of constant $x(t)$. In that case, the gravitational term of equation (3.1) exponentially suppresses the weights of both components of the wave function, but one component is suppressed more strongly than the other, depending on whether $x - \xi$ is positive or negative. One component thus quickly dominates the overall wave function, and the superposition is reduced to just the single product state $\chi_\downarrow \psi^0(x+d)$ or $\chi_\downarrow \psi^0(x-d)$, as is shown in figure 2. That the reduction of the superposed wave function results in just one component in the position basis is a direct consequence of gravity providing the unitarity breaking field in equation (3.1). Because of the special role played by the position variable in the theory of gravity, spatial superpositions become unstable and the pointer basis that defines the possible outcomes of the reduction process consists of spatially localized states. Notice that the spatial reduction of the cantilever wavefunction automatically implies that also the entangled spin state will end up in only one of its components, even though it is not itself subject to any gravitational effects.

The localization dynamics is taken even one step further in the second process induced by the modified Schrödinger equation. Because the single component $\psi^0(x \pm d)$ is itself a wave function spread out in space, it is also subject to the uneven suppression by the gravitational term. As long as the spread of the wave function (i.e. its deBroglie wavelength) is short compared with the distance $\xi \pm d$
A nanoscale experiment

Figure 3. Two typical realizations of the dynamics of equation (3.1) in the presence of a randomly fluctuating $\xi(t)$, starting from the initial wave function $\sqrt{1/4}\chi_0^\uparrow(x + d) + \sqrt{3/4}\chi_0^\downarrow(x - d)$. The fate of the wave function is tracked by the expectation value of the position operator, $\langle \hat{x} \rangle$. The initial wave function is reduced to just one of its components in a time proportional to the inverse of the strength of the gravitational term in equation (3.1). The parameters used in these simulations are the same as in figure 2. (Online version in colour.)

however, the difference in amplification rates of its components is very small, and the secondary reduction of the wave function into a state centred at $x = \xi$ will be very slow (as shown in figure 2).

(b) Born’s Rule

In the presence of a fluctuating $\xi(t)$, the same two processes occur. This time however $x - \xi$ may alternate in time between being positive or negative. It will thus alternately amplify the relative weights of the two different components of the initial superposition, and the question of which component wins out in the end becomes a realization of the ‘gambler’s ruin’ game: each component will randomly increase and decrease in relative weight until one of them completely dominates. Although the fluctuations of the stochastic variable $\xi(t)$ slow down the rise to dominance of either of the components, a final resolution can still be seen to be reached within a typical time scale proportional to the inverse of the gravitational energy $\Delta \equiv G m^2 d^2 / L^3$ defined in equation (3.1) (van Wezel 2010), as expected on general dimensional grounds (Diósi 1987, 1989; Penrose 1996, 1998; van Wezel et al. 2008). This energy scale sets the minimum time scale required by the resonator dynamics to unambiguously display non-unitary effects.

An example of a typical solution of the modified Schrödinger equation with a fluctuating stochastic variable is shown in figure 3. Because $\xi(t)$ is a randomly oscillating variable, it is impossible to predict which of the two components will survive the imposed dynamics. If one of the components has a larger weight in the initial wave function, however, it is more likely to survive. In fact it is possible to mathematically prove that after an infinitely long time, the only possible average outcome of this process is the emergence of Born’s Rule (Zurek 2005; van Wezel 2008). That is, if we repeat the simulation many times always starting from the same initial state $\alpha\chi_0^\uparrow(x + d) + \beta\chi_0^\downarrow(x - d)$, but randomly generating a new $\xi(t)$ for each run, the average fraction of evolutions resulting in the final state $\chi_0^\uparrow\psi^0(x + d)$ converges to $|\alpha|^2$, as in figure 4.
Figure 4. The cumulative proportion of wave functions being reduced to $\chi^+_j \psi^0(x + d)$, starting from $\alpha \chi^+_j \psi^0(x + d) + \beta \chi^+_j \psi^0(x - d)$. The initial dynamics shown in figure 3 was repeated 20000 times for each of three different sets of the parameters $m$ and $d$. For each set the cumulative proportion of outcomes resulting in $\chi^+_j \psi^0(x + d)$ is seen to be $|\alpha|^2$, in accordance with Born’s rule. Here, we used $\alpha = \sqrt{9/20}$. (Online version in colour.)

4. Protocol for the experiment

Returning now to the experimental set-up described in §2, we need to find a way of distinguishing the dynamics of spontaneous unitarity breaking as described by equation (3.1) from both the predictions of alternative models of the quantum to classical crossover, and from the ubiquitous decoherence effects. Especially the latter is not an easy task, because in the proposed set-up, it is not possible to directly observe the time evolution of the state of the cantilever as displayed in figure 3, and we have to rely on ensemble measurements instead.

(a) Distinguishing decoherence from decay

The central observation in the theory of decoherence is that the many unobservable microscopic degrees of freedom, which are necessarily present in any experiment, have the effect, if averaged over an ensemble of measurements, to reduce any initially pure density matrix to a classical mixture of pointer states (Caldeira & Leggett 1983; Joos & Zeh 1985; Prokof’ev & Stamp 2000; Adler 2003; Zurek 2003). Although decoherence and the non-unitary dynamics of equation (3.1) have completely different origins and differ substantially in their descriptions of a single measurement, the ensemble averages look very similar. In terms of the reduced density matrices (i.e. density matrices averaged over all unobservable environmental degrees of freedom), both give rise to an effective, ensemble-averaged evolution of the type

$$\rho(t) = \begin{pmatrix} |\alpha|^2 & \alpha^* \beta e^{-t/\tau} \\ \alpha \beta^* e^{-t/\tau} & |\beta|^2 \end{pmatrix}.$$ (4.1)

Here $\tau$ is either the decoherence time $\tau_{\text{decoh}}$ or the average decay time $\tau_{\text{decay}}$. To distinguish the influence of the non-unitary term from the more usual effects coming from various sources of decoherence, we propose an experimental protocol consisting of two different stages.

Proc. R. Soc. A (2012)
Stage One. In the first instance, the spin and the resonator should be entangled in such a way as to give rise to a spatial superposition of resonator states. This can be achieved in a standard magnetic resonant force microscopy experiment in which the spin is reversed at twice the resonator frequency to drive the resonator, but starting from a superposed spin state. One component of the spin wave function will then enhance the resonator motion more strongly than the other, and as long as coherence can be maintained, the result will be an entangled state in which the resonator motion has a larger amplitude in one component than in the other:

\[
\varphi(t=0) = \alpha \chi_L(x) + \beta \chi_S(x) \]

\[
\Leftrightarrow \rho(t=0) = \begin{pmatrix}
    |\alpha|^2 & \alpha \beta^* \\
    \alpha^* \beta & |\beta|^2
\end{pmatrix},
\]

where \(\psi_L(x)\) and \(\psi_S(x)\) indicate the wave functions of the oscillator with enhanced or suppressed amplitude, respectively, and the spin’s \(S^z\) component in \(\chi_\sigma\) is measured in the co-rotating frame. This state with spatially separated components for the massive resonator will be sensitive to gravity-induced non-unitary effects as well as to decoherence, so that the off-diagonal element of its reduced density matrix after a given time interval \(t_1\) is given by:

\[
\rho_{12}(t_1) = \alpha \beta^* e^{-t_1/\tau_{\text{decoh}}} - t_1/\tau_{\text{decay}}.
\]

The effective dephasing time (which combines the gravitational decay time and the effects of decoherence) can be found after many repetitions of the experiment in which the final state is read out by either coupling the nanomagnet attached to the oscillator to a SQUID through a coil, or alternatively by measuring the fluorescence coming from the spin at the NV centre after the application of a microwave \(\pi/2\) pulse (which rotates the spin orientation into the \(xy\) plane). The effective dephasing time found in this way, in general, depends on both the mass and the shape of the resonator. The characteristics of this dependence can be used as a first means of distinguishing gravity-induced non-unitary dynamics from any given, known source of decoherence. For any given source of decoherence, it is straightforward to work out the functional dependence of \(\tau_{\text{decoh}}\) on the mass of the resonator, its linear size, its geometric shape, etc. In general, one or more of these dependencies will differ from the ones implied by the non-unitary form of equation (3.1). For example, because they depend only on internal degrees of freedom of the oscillator, most sources of decoherence are independent of the angle between the direction of oscillation and the normal to the surface of the gold plate. By measuring the functional dependence of the observed \(\tau\) on such parameters, we can thus rule out most known sources of decoherence. To make sure that no unknown sources of decoherence are at play either, one can then add a second stage to the experiment.

Stage Two. Any source of decoherence, no matter what its precise physical origin is, will arise from the entanglement of the cantilever motion with some other, unobserved degrees of freedom (called the bath). The gravitational decay, on the other hand, only requires the presence of a massive superposition, and takes place even in complete isolation. To differentiate between the two mechanisms, one thus needs to ascertain that the cantilever is not entangled with any bath degrees of freedom, known or unknown. This is done in the next stage, where we...
again start by creating the entangled superposition of equation (4.2), but rather than reading out the amplitude of the resonator after a given time interval, we imprint its state back onto the diamond spin. To do this, one could use a spin echo set-up, and flip the spin at a given instant in time (using an extra microwave $\pi$ pulse, which inverts the orientation of the spin), so that from then on the magnetic force between spin and resonator will tend to damp its motion if it enhanced it before the spin flip, and vice versa. Thus, the effect of the different spin orientations on the oscillator position will be effectively undone. Once the resonator has traced its way back to its original wavefunction, the position of the oscillator state.

The full wave function evolution in the presence of both decay and bath degrees of freedom can thus schematically be written as:

$$
\begin{align*}
\varphi(t = 0) &= [\alpha\chi_\uparrow + \beta\chi_\downarrow]\psi^0(x), \\
\varphi(0 < t < t_1) &= \alpha\chi_\downarrow\psi^L(x, t) + \beta\chi_\uparrow\psi^S(x, t), \\
\varphi(t_1 < t < t_2) &= \alpha\chi_\downarrow\psi^L(x, t) + \beta\chi_\uparrow\psi^S(x, t)
\end{align*}
$$

and

$$
\varphi(t = t_2) = [\alpha\chi_\downarrow + \beta\chi_\uparrow]\psi^0(x).
$$

Here, the spin echo pulse was applied at $t = t_1$, and the time dependence of $\psi^{L,S}(x, t)$ signifies a growing (or shrinking) oscillation amplitude as a function of time for $t < t_1$, but the opposite for $t > t_1$, owing to the opposing influence of the flipped spin. If a non-unitary term is present, the components $\alpha$ and $\beta$ become time dependent, but the oscillator state will still be brought back to $\psi^0(x)$ at $t = t_2$. Likewise, decoherence may result from the entanglement of these states with bath degrees of freedom, but these do not alter the final oscillator state.

After decoupling the spin state and the oscillator state at $t_2$, there will not be any additional gravitational decay because there is no longer any spatially superposed mass. Any decoherence caused by the entanglement of the diamond spin with unknown internal (bath) degrees of freedom of the resonator however, will continue unabated even if the resonator remains at rest. After all, owing to the rigidity of the resonator (and the associated Goldstone theorem), its internal degrees of freedom are decoupled from its collective coordinates. Even though the spin flip at $t = t_1$ effectively reverses the time evolution of the collective, centre-of-mass motion of the resonator, the dynamics of the internal, bath degrees of freedom is not reversed and they remain in an entangled state even after $t = t_2$. The full wave function evolution in the presence of both decay and bath degrees of freedom can thus schematically be written as:

$$
\begin{align*}
\varphi(t = 0) &= [\alpha(0)\chi_\uparrow + \beta(0)\chi_\downarrow]\psi^0(x)\phi^0(0), \\
\varphi(0 < t < t_1) &= \alpha(t)\chi_\downarrow\psi^L(x, t)\phi^L(t) + \beta(t)\chi_\uparrow\psi^S(x, t)\phi^S(t), \\
\varphi(t_1 < t < t_2) &= \alpha(t)\chi_\downarrow\psi^L(x, t)\phi^L(t) + \beta(t)\chi_\uparrow\psi^S(x, t)\phi^S(t)
\end{align*}
$$

and

$$
\varphi(t > t_2) = [\alpha(t_2)\chi_\downarrow\phi^L(t) + \beta(t_2)\chi_\uparrow\phi^S(t)]\psi^0(x).
$$

**Proc. R. Soc. A (2012)**
Here $\phi^{L,S}(t)$ represent the evolution of the bath degrees of freedom, as influenced by the different amplitude states of the oscillator. Their time evolution continues beyond $t = t_2$, while the components $\alpha$ and $\beta$ become time-independent after the centre-of-mass state has returned to $\psi^0(x)$. After tracing out the bath degrees of freedom and simultaneously averaging over many realizations of the stochastic dynamics imposed by the non-unitary time evolution, the off-diagonal element of the reduced density matrix for the final state yields:

$$\rho_{12}(t > t_2) = \alpha(0)\beta(0)^*e^{-t_2(1/\tau_{\text{decoh}} + 1/\tau_{\text{decay}})}e^{-(t-t_2)/\tau_{\text{decoh}}}.$$ (4.6)

At the end of the second experimental stage then, the coherence of the diamond spin can again be monitored via optical readings in many repetitions of the experiment. The initial suppression of the off-diagonal matrix elements by the factor $e^{-t_2/\tau_{\text{decoh}}}$ is known from the first stage of the experiment. The remaining suppression of the off-diagonal matrix elements by the factor $e^{-(t-t_2)/\tau_{\text{decoh}}}$ can be attributed purely to sources of decoherence. It will most probably be dominated by the well-characterized decoherence of an isolated diamond spin owing to sources within the diamond environment itself. For the experimental parameters proposed here, this time scale should be in the millisecond range, and can be easily recognized. Any remaining sources of decoherence, which had not been identified and eliminated in the first experimental stage can now be recognized as an apparent contribution to $e^{-(t-t_2)/\tau_{\text{decoh}}}$, in excess of that expected from the known, pristine diamond environment. Once these remaining unidentified bath degrees of freedom have been eliminated (i.e. through additional cooling), the dephasing measured in the first stage of the experiment can be conclusively attributed to a decay process instead of decoherence.

(b) Distinguishing different decay models

After ruling out decoherence as the source for the suppression of off-diagonal matrix elements, we still need to ascertain whether the observed effective decay time $\tau$ is owing to the process of spontaneous unitarity breaking, which we focused on in this paper, or to some other non-unitary mechanism. The most well-known alternatives of this kind fall in the Ghirardi–Rimini–Weber (GRW) and continuous spontaneous localization (CSL) classes (Pearle 1976; Ghirardi et al. 1986; Pearle & Squires 1994; Bassi & Ghirardi 2003). Other proposed scenarios for the quantum to classical crossover, such as hidden variable and many worlds theories are fully unitary (de Broglie 1928; Bohm 1952a,b; Everett 1957; Deutsch 1999), and the observation of any non-zero decay time in the proposed experiment would thus suffice to rule out their relevance in the description of the resonator dynamics.

The main assumption in both the CSL and the GRW theories is the existence of a specific process beyond the realm of applicability of Schrödinger’s equation. In the GRW theory, this process is assumed to be the spontaneous and instantaneous spatial localization of an elementary particle at particular intervals in time (Ghirardi et al. 1986). If the frequency with which these localization events occur is low enough, it will take an immeasurably long time (on average) for any individual particle to undergo such an event. Within an extended object consisting of a macroscopic number of particles on the other hand, it will not
take long before at least one of the particles is localized. Assuming that the extended object possesses some degree of rigidity, the localization of a single particle within it will suffice to give the entire wave function a definite position in space. Macroscopic objects are thus rapidly reduced to position eigenstates, while microscopic particles are free to spread out in space. Born’s rule can be built into this construction by making additional assumption about the localization process (Ghirardi et al. 1986).

Because the localization process proposed by GRW acts on individual microscopic particles, and does not depend on the overall state of the system, its predicted decay time scales only with the number of particles involved in the massive superposition. A dependence of the decay time $\tau_{\text{decay}}$ in equation (4.6) on the involved mass or shape of the resonator will thus clearly indicate a departure from the predictions of the GRW theory, and require additional non-unitary effects.

The CSL models can be seen as extensions of the original GRW model in which the addition to quantum mechanics is no longer a set of instantaneous localization events, but rather a continuous process that constantly acts to gradually localize the individual particles. In most modern CSL models, a smeared mass density is taken to be the variable that specifies both the rate and the final states of the reduction process (Pearle 1976, 2009; Pearle & Squires 1994; Bassi & Ghirardi 2003). The obtained average rate of localization ensures that microscopic particles take an immeasurably long time to localize, while macroscopic objects again localize almost instantaneously owing to their rigidity. Born’s rule can emerge spontaneously if the localization process is stochastic, which is modelled by the inclusion of white noise in the dynamical equations (Bassi & Ghirardi 2003).

The final form of the reduction process described by the CSL model resembles the dynamics of spontaneous unitarity breaking in equation (3.1) in various ways. Both predict a reduction of the superposed initial state within a time scale proportional to the square of the total mass of the involved object; both involve a random variable which ensures that Born’s rule can be obeyed; and both mechanisms select the position basis as the pointer basis to which macroscopic objects must be reduced.

However, there are also clear differences between the two approaches. For example, as discussed in the review articles of Bassi & Ghirardi (2003), and of Pearle (2009), the effective reduction rate in the CSL model is set directly by the spread of the mass density distributions in different components of the initial wave function, independent of the relative locations of their centres of mass (if the distributions are initially well separated). Thus, in contrast to equation (3.1), the reduction rate of a macroscopic superposition is independent of the relative distance between the massive bodies in its components, as long as they are initially non-overlapping. Such a difference would be obvious in the experiment, if we could construct a superposition of the massive plate over distances greater than its thickness. If on the other hand, the two components of the resonator wave function do have a substantial spatial overlap throughout the experiment, we need to rely on the dependence of the decay time on the precise geometry of the massive plate to differentiate between the predictions of the CSL model and spontaneous unitarity breaking. In the CSL model, the effective decay time can be shown to be proportional to $d/L^2$, where $d$ is the thickness of the non-overlapping part of
the involved mass distributions (Bassi & Ghirardi 2003). Instead, spontaneous unitarity breaking can be seen from equation (3.1), to predict the rate to be proportional to \( d^2/L^3 \) (van Wezel et al. 2008).

After ruling out decoherence as the cause for the decay of the off-diagonal matrix element in the two-stage protocol, we can thus use the dependence of the observed decay time on the mass and geometry of the proposed set-up to conclusively distinguish the effects of all currently proposed descriptions of the quantum to classical crossover.

5. Conclusions

We have proposed a nanoscale experiment, realizable with present-day technology, which can be used to produce an entangled state of a NV spin in diamond and a mechanical resonator. Because the entangled state involves the superposition of a massive cantilever over a sizable distance, we are forced to consider the effects that a quantum to classical crossover might have on the dynamics of this state. The proposed experiment is focused on testing one particular scenario for this crossover, which describes the disappearance of unitary quantum dynamics in the thermodynamic limit as a form of spontaneous symmetry breaking, akin to the spontaneous symmetry breaking observed in macroscopic crystals, magnets, etc. By constructing an effective, phenomenological description of the proposed experiment, it is shown that the predicted effects of spontaneous unitarity breaking on the time evolution of the resonator state can be experimentally tested. A specific experimental protocol is proposed in order to distinguish the effects of spontaneous unitarity breaking from those of both decoherence and other proposed models for the quantum to classical crossover.

The phenomenological description of the resonator dynamics predicts the reduction of a superposed initial state to just one of its components to take place on a time scale which lies just within experimental reach. The reduction of the cantilever state also implies the reduction of the entangled spin state, even though the spin itself is not directly under the influence of the broken unitarity. The emergence of both a pointer basis and Born’s rule for the reduction process can be seen to be a direct consequence of the influence of gravity on quantum mechanics.

One of the authors (J.v.W.) acknowledges support by the US DOE, Office of Science, under contract no. DE-AC02-06CH11357, for the part of the work done at Argonne National Laboratory.

Appendix A. The norm of the wave function

To see how the usual probabilistic predictions of quantum mechanics may emerge from the modified time evolution of equation (3.1), even though it does not conserve the normalization of the wave function (van Wezel 2010), consider a particular run of the proposed experiment, with the spin of the diamond NV centre initially in a superposition of \( S^z \) states, while the resonator is in a single position state:

\[
\varphi(t < 0) = [\alpha \chi_\uparrow + \beta \chi_\downarrow] \psi^0(x). \tag{A1}
\]
Here, without loss of generality, we assume $|\alpha|^2 + |\beta|^2 = 1$. This state is a stable state, in the sense that the diamond NV centre is not affected by the non-unitary term of the modified Schrödinger equation as it does not involve a massive superposition, while the oscillator is in a state with a single, well-defined position, and is thus subject only to the extremely slow dynamics indicated by the dashed curves in figure 2. For the estimated experimental parameters discussed in §2, the slow motion of the oscillator happens on an unmeasurably long time scale. If we were to consider even heavier objects (like, for example, the pointer on a regular measurement machine), this time scale becomes even longer. For all practical purposes, the product state $\varphi(t < 0)$ is thus a static configuration, with a conserved norm.

In the initial phase of the experiment, the spin state and the oscillator wave function become entangled by the coupling of the spin orientation to the deflection of the cantilever in the usual MFRM set-up. If this can be done within a time that is short compared with both the intrinsic coherence time of the spin and the typical decay time imposed by the non-unitary field, the resulting state will be given by:

$$\varphi(t = 0) = \alpha \chi_1 \psi^0(x + d) + \beta \chi_1 \psi^0(x - d). \quad (A2)$$

This state is an entangled state, which involves the superposition of a massive object over different locations. It is therefore subject also to the faster non-unitary decay dynamics (indicated by solid curves in figure 2). For the experimental parameters proposed in §2, the fast non-unitary dynamics takes milliseconds to complete, and is within the measurable regime. For heavier objects, this time scale quickly becomes unmeasurably short. If we were to consider a truly macroscopic superposition (involving, for example, the pointer on a regular measurement device), the decay would, for all practical purposes, be instantaneous.

During the decay dynamics, the norm of the wave function is not conserved. The result, if the decay happens on a time scale short compared with the involved intrinsic decoherence times, will be either one of two possible outcomes:

$$\varphi(t \gg 0) = \begin{cases} N \chi_1 \psi^0(x + d) \\ \text{or} \\ N \chi_1 \psi^0(x - d), \end{cases} \quad (A3)$$

where the norm $N$ depends on the precise dynamics, and is certainly not equal to one. Despite the lack of a normalized wave function, there can be no doubt about the physical interpretation of the final states in equation (A3). Each of the possible outcomes is a single product state, which combines an eigenfunction of the $z$-projection operator of the spin in the diamond NV centre with a single, well-defined value for the spatial location of the oscillator. It must thus be interpreted as representing a spin that is fully oriented along the $z$-axis, and an oscillator in a (classical or generalized coherent) state centred at position $x \pm d$. The position of the oscillator can be registered and recorded, and thus effectively serves as a measurement of the spin state. Notice that the final state is once again a stable configuration, which is insensitive to further non-unitary influences. As long as no sufficiently massive superpositions are created, its further time evolution will be unitary for all practical purposes, and its norm $N$ will be conserved.

Proc. R. Soc. A (2012)
We can repeat the above experiment many times, and each time record which of the two states in equation (A3) is the outcome of that particular realization. The number of experiments resulting in a state with the oscillator centred at \( x + d \) divided by the total number of conducted experiments will then be seen to converge to the value \(|a|^2\), as indicated in figure 4. Since it is impossible to predict which of the two outcomes will be realized in any one particular experiment, this is equivalent to having a probabilistic reduction of the initial state to just one of its components, with the probability for obtaining a given component set by its squared weight in the initial wave function:

\[
|\alpha^2| \quad \text{and} \quad |\beta|^2.
\]

As noted before, the reduction process would happen unmeasurably fast if we were to employ any of the truly massive devices that are customarily used as quantum measurements machines. In that case, the dynamics described here thus reproduces the instantaneous ‘collapse of the wave function’ that is traditionally postulated as an addition to Schrödinger’s equation. For devices much lighter than the proposed micromechanical oscillator on the other hand, the non-unitary reduction process takes an unmeasurably long time to complete, and the dynamics remains effectively unitary, as prescribed by Schrödinger’s equation. Both the unitary quantum dynamics of microscopic particles and the non-unitary behaviour of classical objects thus emerge naturally from the dynamics of equation (3.1). In particular, the usual probabilistic predictions of quantum mechanics are recovered, in spite of the absence of a conserved norm of the wave function.

Appendix B. The perturbation to Schrödinger’s equation

The dynamical Schrödinger equation can be seen to be subject to spontaneous symmetry breaking, in the sense that its result may be qualitatively affected in the thermodynamic limit by only an infinitesimally weak unitarity breaking field. This description of broken unitarity is a straightforward extension of the well-known mechanism of spontaneous symmetry breaking in crystals, magnets, and so on \( \text{van Wezel (2010)} \). For such spontaneous symmetry breaking to be effective in large, but finite, objects, there must exist a small, but finite, non-unitary contribution to its dynamics. In this appendix, we give a brief sketch of how a non-unitary term such as the one in equation (3.1) may be seen to naturally arise from considerations of the interplay between quantum mechanics and general relativity (\text{van Wezel 2010}).

As it turns out, the fundamental building blocks of these theories (unitarity for quantum mechanics and general covariance for general relativity) are mutually incompatible concepts. All theories trying to bridge the gap between general relativity and quantum mechanics have to abandon either one or both of these principles. For example, the background dependence of non-equilibrium string

\[1\text{Notice that for a general initial norm } |N_0|^2 = |a|^2 + |\beta|^2, \text{ we find probabilities } |a|^2/|N_0|^2 \text{ and } |\beta|^2/|N_0|^2 \text{ instead.}\]
J. van Wezel and T. H. Oosterkamp

theory breaks general covariance, while an equilibrium formulation in Euclidean spacetime is necessarily dissipative and thus non-unitary. The conflict can be made apparent by comparing simple examples of time evolution in the two theories (Penrose 1996, 1998). For the purpose of this paper, we remain agnostic about the ultimate role of general covariance, but we assume that unitarity is not an exact symmetry of nature, and may thus be spontaneously broken (van Wezel 2010).

The density of the resonator in our experiment is rather modest compared with the densities occurring in black holes or in high energy physics. The dynamics of the resonator will, therefore, be very close to purely quantum mechanical, and we expect to be able to include the effects of general relativity as minor perturbations to Schrödinger’s equation. Rather than identifying the operator \( \frac{d}{dt} \) with just \( -\frac{i}{\hbar} \hat{H} \), we therefore introduce a small perturbative term \( \hat{H}' \):

\[
i\hbar \frac{d}{dt} \psi(\vec{r}, t, \sigma) = [\hat{H} + \hat{H}'] \psi(\vec{r}, t, \sigma),
\]

where the perturbation \( \hat{H}' \) is assumed to be due to the influence of general relativity.

To find the functional form of the perturbation, we start by assuming that the defining property of the conflict between general relativity and quantum mechanics is the essential non-unitarity of the former. Other ingredients (such as a possible non-linearity) will be left to higher order terms. The correction term \( \hat{H}' \) then cannot be a Hermitian operator, but on general grounds, we would like the time evolution to be invertible even in the presence of gravity, so that physics makes sense if time flows backwards as well as forwards (although it does not necessarily have to look the same). One possible way to enforce this constraint is by writing

\[
i \hbar \frac{d}{dt} \psi = -\frac{i}{\hbar} (\hat{H} - i\hat{X}) \psi,
\]

where \( \hat{X} \) is a linear and Hermitian operator. The fact that the time evolution generated in this way does not conserve energy (as measured by \( \hat{H} \)) agrees with the lack of a locally conserved energy concept in a non-static configuration of general relativity. Of course globally, energy should be a conserved quantity, and we will have to choose \( \hat{X} \) such that it takes account of that restriction. In fact, the presence of an order parameter in rigid macroscopic objects can be shown to automatically restore global energy conservation in the thermodynamic limit (van Wezel 2010).

The second step in our argumentation is to relate the ‘energy’ term \( \hat{X} \) to a measure for the extent to which the presence of a given quantum superposition is in conflict with the requirements of general relativity, and thus gives rise to a non-unitary correction to quantum mechanics. It has been shown by Diósi (1987, 1989), and independently by Penrose (1996, 1998) that there exists a covariant way of constructing such a measure. For an equal superposition of two different mass distributions \( \rho_1 \) and \( \rho_2 \) in the Newtonian limit, they found this measure to
be (Diósi 1987, 1989; Penrose 1996, 1998):

\[
\Delta = -4\pi G \int \left[ \frac{\rho_1(x) - \rho_2(x)}{|x - y|} \right] \frac{\rho_1(y) - \rho_2(y)}{|x - y|} d^3x d^3y. \tag{B.3}
\]

In the special case in which the superposed mass \( m \) is a block, which is evenly superposed over a distance \( x \) small compared with its width \( L \) (as in figure 5), the integrals of equation (B.3) can be evaluated to yield (van Wezel et al. 2008)

\[
\Delta = G \frac{m^2}{2L^3} x^2. \tag{B.4}
\]

The form of equation (B.4) allows a straightforward generalization to the case of a generic superposition, which consists of any number of components carrying arbitrary weights in the wave function. If \( \hat{x} \) is the standard quantum mechanical operator measuring the position of the block’s centre of mass (with the zero of position at the overall centre of mass of the initial wave function), then we can interpret \( \Delta \) as the expectation value of the quantum operator

\[
\hat{\Delta} = G \frac{m^2}{2L^3} \hat{x}^2. \tag{B.5}
\]

Notice that although we started with a semi-classical definition of the incompatibility measure in equation (B.3), the final expression of equation (B.5) is a fully quantum mechanical, Hermitian operator. This quantized form is only applicable to the specific case of a massive block superposed over small distances along one spatial direction, and in that sense is not as general as the original semiclassical expression. However, because it is an operator form, it can be applied to any such wave function of the massive block, and it is not restricted to an even distribution over just two states. For such a more general wave function, the expectation value of the operator is still a good measure of the uncertainty introduced into the concept of a locally conserved energy by the application of general covariance. These properties of the operator form of \( \hat{\Delta} \) suggest that it may also be an appropriate form for the first-order correction to Schrödinger’s time evolution in equation (B.2):

\[
\hat{X} \propto G \frac{m^2}{2L^3} \hat{x}^2. \tag{B.6}
\]
Finally, we need to address the fact that the spatial distance between points in separate space–times is an ill-defined notion in general relativity (Penrose 1996, 1998). In this regard, the expression of equation (B6) for the first-order correction is not fully satisfactory. The operator \( \hat{x} \) measures the position of the centre of mass in a component of the massive superposition using a universal coordinate system, which is applied equally to all other components. We needed to define such a ‘best possible’ match between the coordinate systems of different components to first calculate the semi-classical measure for its inappropriateness (equation (B3)), but the final form of the correction to Schrödinger’s equation should not depend on any such (arbitrary) specific choice of coordinate system. To model the ill-definedness in the definition of \( \hat{x} \), we introduce a random variable \( \xi \), and replace the measure of distance by \((\hat{x} - \xi)\):

\[
\hat{X} = G \frac{m^2}{2L^3}[\hat{x} - \xi]^2. \tag{B7}
\]

As emphasized also in the main text, the introduction of a random variable here should be seen only as a poor man’s approach towards modelling an essentially ill-defined quantity. We do not expect the full theory of quantum gravity to be non-deterministic, but we are forced to take into account an effectively random correction to the notion of position, if we are to consider gravity’s first-order perturbation to quantum dynamics. The equality of the numerical pre-factors of \( \hat{x} \) and \( \xi \) reflects the fact that the correction owing to gravity should depend on the spatial spread of the wave function.

We thus finally reproduce equation (3.1) as the effective, phenomenological description of the minimal correction to the unitary Schrödinger equation owing to the influence of general covariance (van Wezel 2010):

\[
\frac{d}{dt} \psi = -\frac{i}{\hbar} \left( \hat{H} - iG \frac{m^2}{2L^3}[\hat{x} - \xi]^2 \right) \psi, \tag{B 8}
\]

This final form of the modified Schrödinger equation can be straightforwardly integrated using standard numerical methods to yield a phenomenological prediction for the expected time evolution of the proposed resonator experiment.

References

Adler, S. L. 2003 Why decoherence has not solved the measurement problem: a response to P. W. Anderson. Stud. Hist. Phil. Mod. Phys. 34 135–142. (doi:10.1016/S1355-2198(02)00086-2)

Anderson, P. W. 1972 More is different. Science 177, 393–396. (doi:10.1126/science.177.4047.393)

Balasubramanian, G. et al. 2009 Ultralong spin coherence time in isotopically engineered diamond. Nat. Mater. 8, 383–387. (doi:10.1038/nmat2420)

Bassi, A. & Ghirardi, G. C. 2003 Dynamical reduction models. Phys. Repts. 379, 257–426. (doi:10.1016/S0370-1573(03)00103-0)

Bohm, D. 1952a A suggested interpretation of the quantum theory in terms of ‘Hidden’ variables. I. Phys. Rev. 85, 166–179. (doi:10.1103/PhysRev.85.166)

Bohm, D. 1952b A suggested interpretation of the quantum theory in terms of ‘Hidden’ variables. II. Phys. Rev. 85, 180–193. (doi:10.1103/PhysRev.85.180)

de Broglie, L. 1928 Électrons et photons: rapports et discussions du cinquième conseil de physique (ed. J. Bordet). Paris, France: Gauthier-Villars.
Caldeira, A. O. & Leggett, A. J. 1983 Path integral approach to quantum Brownian motion. *Physica A* **121**, 587–616. (doi:10.1016/0378-4371(83)90013-4)

Cho, A. 2003 Researchers race to put the quantum into mechanics. *Science* **299**, 36–37. (doi:10.1126/science.299.5603.36)

Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T. & Rugar, D. 2009 Nanoscale magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **106**, 1313–1317. (doi:10.1073/pnas.0812068106)

Deutsch, D. 1999 Quantum theory of probability and decisions. *Proc. R. Soc. Lond. A* **455**, 3129–3137. (doi:10.1098/rspa.1999.0443)

Diósi, L. 1984 Gravitation and quantum-mechanical localization of macro-objects. *Phys. Lett. A* **105**, 199–202. (doi:10.1016/0375-9601(84)90397-9)

Diósi, L. 1987 A universal master equation for the gravitational violation of quantum mechanics. *Phys. Lett. A* **120**, 377–381. (doi:10.1016/0375-9601(87)90681-5)

Diósi, L. 1989 Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev. A* **40**, 1165–1174. (doi:10.1103/PhysRevA.40.1165)

Everett, H. 1957 ‘Relative state’ formulation of quantum mechanics. *Rev. Mod. Phys.* **29**, 454–462. (doi:10.1103/RevModPhys.29.454)

Ghirardi, G. C., Rimini, A. & Weber, T. 1986 Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D* **34**, 470–491. (doi:10.1103/PhysRevD.34.470)

Huttel, A. K., Steele, G. A., Witkamp, B., Poot, M., Kouwenhoven, L. P. & van der Zant, H. S. J. 2009 Carbon nanotubes as ultrahigh quality factor mechanical resonators. *Nano Lett.* **9**, 2547–2552. (doi:10.1021/nl900612h)

Joos, E. & Zeh, H. D. 1985 The emergence of classical properties through interaction with the environment. *Z. Phys. B Cond. matter* **59**, 223–243. (doi:10.1007/BF01725541)

Kleckner, D., Pikovski, I., Jeffrey, E., Ament, L., Eliel, E., van den Brink, J. & Bouwmeester, D. 2008 Creating and verifying a quantum superposition in a micro-optomechanical system. *New J. Phys.* **10**, 095020. (doi:10.1088/1367-2630/10/9/095020)

LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. 2004 Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77. (doi:10.1126/science.1094419)

LaHaye, M. D., Suh, J., Echternach, P. M., Schwab, K. C. & Roukes, M. L. 2009 Nanomechanical measurements of a superconducting qubit. *Nature* **459**, 960–964. (doi:10.1038/nature08093)

Li, T., Kheifets, S. & Raizen, M. G. 2011 Millikelvin cooling of an optically trapped microsphere in vacuum. *Nat. Phys.* **7**, 527–530. (doi:10.1038/nphys1952)

Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. 2003 Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401. (doi:10.1103/PhysRevLett.91.130401)

Moroz, I., Penrose, R. & Tod, P. 1998 Spherically-symmetric solutions of the Schrödinger–Newton equations. *Class. Quant. Gravity* **15**, 2733. (doi:10.1088/0264-9381/15/9/019)

Müller-Ebhardt, H., Rehbein, H., Schnabel, R., Danzmann, K. & Chen, Y. 2008 Entanglement of macroscopic test masses and the standard quantum limit in laser interferometry. *Phys. Rev. Lett.* **100**, 013601. (doi:10.1103/PhysRevLett.100.013601)

Müller-Ebhardt, H., Rehbein, H., Li, C., Mino, Y., Somiya, K., Schnabel, R., Danzmann, K. & Chen, Y. 2009 Quantum-state preparation and macroscopic entanglement in gravitational-wave detectors. *Phys. Rev. A* **80**, 043802. (doi:10.1103/PhysRevA.80.043802)

Naik, A., Buu, O., LaHaye, M. D., Armour, A. D., Clerk, A. A., Blencowe, M. P. & Schwab, K. C. 2006 Cooling a nanomechanical resonator with quantum back-action. *Nature* **443**, 193–196. (doi:10.1038/nature05027)

O’Connell, A. D. *et al.* 2010 Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703. (doi:10.1038/nature08967)

Pearle, P. 1976 Reduction of the state vector by a nonlinear Schrödinger equation. *Phys. Rev. D* **13**, 857–868. (doi:10.1103/PhysRevD.13.857)

Pearle, P. 2009 How stands collapse II. In *Quantum reality, relativistic causality and closing the epistemic circle: essays in honour of Abner Shimony* (eds W. Myrvold & J. Christian). The Western Ontario Series in Philosophy of Science, pp. 257–292. Berlin, Germany: Springer.

Pearle, P. & Squires, E. 1994 Bound state excitation, nucleon decay experiments, and models of wave function collapse. *Phys. Rev. Lett.* **73**, 1–5. (doi:10.1103/PhysRevLett.73.1)
Penrose, R. 1996 On gravity’s role in state vector reduction. *Gen. Rel. Gravity* **28**, 581–600. (doi:10.1007/BF02105068)

Penrose, R. 1998 Quantum computation, entanglement and state reduction. *Phil. Trans. R. Soc. Lond. A* **356**, 1927–1939. (doi:10.1098/rsta.1998.0256)

Prokof’ev, N. V. & Stamp, P. C. E. 2000 Theory of the spin bath. *Rep. Prog. Phys.* **63**, 669. (doi:10.1088/0034-4885/63/4/204)

Rabl, P., Cappellaro, P., Gurudev Dutt, M. V., Jiang, L., Maze, J. R. & Lukin, M. D. 2009 Strong magnetic coupling between an electronic spin qubit and a mechanical resonator. *Phys. Rev. B* **79**, 041302. (doi:10.1103/PhysRevB.79.041302)

Remus, L. G., Blencowe, M. P. & Tanaka, Y. 2009 Damping and decoherence of a nanomechanical resonator due to a few two-level systems. *Phys. Rev. B* **80**, 174103. (doi:10.1103/PhysRevB.80.174103)

Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. 2004 Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332. (doi:10.1038/nature02658)

Takahashi, S., Hanson, R., van Tol, J., Sherwin, M. S. & Awschalom, D. D. 2008 Quenching spin decoherence in diamond through spin bath polarization *Phys. Rev. Lett.* **101**, 047601. (doi:10.1103/PhysRevLett.101.047601)

Unruh, W. G. & Zurek, W. H. 1989 Reduction of a wave packet in quantum Brownian motion. *Phys. Rev. D* **40**, 1071–1094. (doi:10.1103/PhysRevD.40.1071)

Vinante, A., Schinkelshoek, L., Usenko, O., Wijts, G. & Oosterkamp, T. H. 2011 Magnetic resonance force microscopy of paramagnetic electron spins at millikelvin temperatures. (http://arxiv.org/abs/1105.3395).

van Wezel, J. 2008 Quantum dynamics in the thermodynamic limit. *Phys. Rev. B* **78**, 054301. (doi:10.1103/PhysRevB.78.054301)

van Wezel, J. 2010 Broken time translation symmetry as a model for quantum state reduction. *Symmetry* **2**, 582–608. (doi: 10.3390/sym2020582) (open access).

van Wezel, J. & van den Brink, J. 2008 The Schrödinger-Newton equation as a possible generator of quantum state reduction. *Phil. Mag.* **88**, 1659–1671. (doi:10.1080/14786430802251439)

van Wezel, J., Oosterkamp, T. H. & Zaanen, J. 2008 Towards an experimental test of gravity-induced quantum state reduction. *Phil. Mag.* **88**, 1005–1026. (doi:10.1080/14786430801941824)

Zurek, W. H. 1991 Decoherence and the transition from quantum to classical. *Phys. Today* **44**, 36. (doi:10.1063/1.881293)

Zurek, W. H. 2003 Decoherence, einselection and the quantum origins of the classical. *Rev. Mod. Phys.* **75**, 715–775. (doi:10.1103/RevModPhys.75.715)

Zurek, W. H. 2005 Probabilities from entanglement, Born’s rule $p_k = |\psi_k|^2$ from envariance. *Phys. Rev. A* **71**, 052105. (doi:10.1103/PhysRevA.71.052105)