Does gravity induce wavefunction collapse? An examination of Penrose's argument

Shan Gao*

January 20, 2012

Abstract

According to Penrose, the fundamental conflict between the superposition principle of quantum mechanics and the general covariance principle of general relativity entails the existence of wavefunction collapse, e.g. a quantum superposition of two different space-time geometries will collapse to one of them due to the ill-definedness of the time-translation operator for the superposition. In this paper, we argue that Penrose's conjecture on gravity's role in wavefunction collapse is debatable. First of all, it is still a controversial issue what the exact nature of the conflict is and how to resolve it. Secondly, Penrose's argument by analogy is too weak to establish a necessary connection between wavefunction collapse and the conflict as understood by him. Thirdly, the conflict does not necessarily lead to wavefunction collapse. For the conflict or the problem of ill-definedness for a superposition of different space-time geometries also needs to be solved before the collapse of the superposition finishes, and once the conflict has been resolved, the wavefunction collapse will lose its physical basis relating to the conflict. In addition, we argue that Penrose's suggestions for the collapse time formula and collapse states are also problematic.

In standard quantum mechanics, it is postulated that when the wave function of a quantum system is measured by a macroscopic device, it no longer follows the linear Schrödinger equation, but instantaneously collapses to one of the wave functions that correspond to definite measurement results. However, this collapse postulate is not satisfactory, as it does not explain why and how the wave function collapses during a measurement. There have been various conjectures on the origin of wavefunction collapse, and the most promising one is Penrose's gravity-induced collapse argument (Penrose 1996). In this paper, we will present a critical analysis of Penrose's trenchant argument.

It seems very natural to guess the collapse of the wave function is induced by gravity. The reasons include: (1) gravity is the only universal force being present in all physical interactions; (2) gravitational effects grow with the size of the objects concerned, and it is in the context of macroscopic objects that linear superpositions may be violated. The gravity-induced collapse conjecture can

*Unit for HPS and Centre for Time, University of Sydney, NSW 2006, Australia. E-mail: sgao7319@uni.sydney.edu.au.

be traced back to Feynman (1995)[1]. In his *Lectures on Gravitation*, Feynman considered the philosophical problems in quantizing macroscopic objects and contemplates on a possible breakdown of quantum theory. He said, "I would like to suggest that it is possible that quantum mechanics fails at large distances and for large objects, it is not inconsistent with what we do know. If this failure of quantum mechanics is connected with gravity, we might speculatively expect this to happen for masses such that $GM^2/\hbar c = 1$, of M near 10^{-5} grams."

Partly inspired by Feynman's suggestion, Penrose proposed a concrete gravity-induced collapse argument (Penrose 1996). The argument is based on a fundamental conflict between the superposition principle of quantum mechanics and the general covariance principle of general relativity. The conflict can be seen by considering the superposition state of a static mass distribution in two different locations, say position A and position B. On the one hand, according to quantum mechanics, the valid definition of such a superposition requires the existence of a definite space-time background, in which position A and position B can be distinguished. On the other hand, according to general relativity, the space-time geometry, including the distinguishability of position A and position B, cannot be predetermined, and must be dynamically determined by the position superposition state. Since the different position states in the superposition determine different space-time geometries, the space-time geometry determined by the whole superposition state is indefinite, and as a result, the superposition and its evolution cannot be consistently defined. In particular, the definition of the time-translation operator for the superposed space-time geometries involves an inherent ill-definedness, and this leads to an essential uncertainty in the energy of the superposed state. Then by analogy Penrose argued that this superposition, like an unstable particle in usual quantum mechanics, is also unstable, and it will decay or collapse into one of the two states in the superposition after a finite lifetime. Moreover, Penrose suggested that the essential energy uncertainty in the Newtonian limit is proportional to the gravitational self-energy E_Δ of the difference between the two mass distributions, and the collapse time, analogous to the half-life of an unstable particle, is

$$T \approx \hbar / E_\Delta \tag{1}$$

This criterion is very close to that put forward by Diósi (1989) earlier, and it is usually called the Diósi-Penrose criterion. Later, Penrose (1998) further suggested that the collapse states are the stationary solutions of the Schrödinger-Newton equation.

Now let's examine Penrose's gravity-induced collapse argument in detail. The crux of the argument is whether the conflict between quantum mechanics and general relativity requires that a quantum superposition of two space-time geometries must collapse after a finite time. We will argue in the following that the answer is negative. First of all, although it is widely acknowledged that there exists a fundamental conflict between the superposition principle of quantum mechanics and the general covariance principle of general relativity, it is still a controversial issue what the exact nature of the conflict is and how to solve it. For example, it is possible that the conflict may be solved by

[1]It is worth noting that Feynman considered this conjecture even earlier at the 1957 Chapel Hill conference (DeWitt and Rickles 2011, ch.22).

reformulating quantum mechanics in a way that does not rely on a definite spacetime background (see, e.g. Rovelli 2011).

Secondly, Penrose's argument by analogy seems too weak to establish a necessary connection between wavefunction collapse and the conflict between general relativity and quantum mechanics. Even though there is an essential uncertainty in the energy of the superposition of different space-time geometries, this kind of energy uncertainty is different in nature from the energy uncertainty of unstable particles or unstable states in usual quantum mechanics (Gao 2010). The former results from the ill-definedness of the time-translation operator for the superposed space-time geometries (and its nature seems still unclear), while the latter exists in a definite spacetime background, and there is a well-defined time-translation operator for the unstable states. Moreover, the decay of an unstable state (e.g. an excited state of an atom) is a natural result of the linear Schrödinger evolution, and the process is not random but deterministic. In particular, the decay process is not spontaneous but caused by the background field constantly interacting with the unstable state, e.g. the state may not decay at all when in a very special background field with bandgap (Yablonovitch 1987). By contrast, the hypothetical decay or collapse of the superposed space-time geometries is spontaneous, nonlinear and random. In short, there exists no convincing analogy between a superposition of different space-time geometries and an unstable state in usual quantum mechanics. Accordingly, one cannot argue for the collapse of the superposition of different space-time geometries by this analogy. Although an unstable state in quantum mechanics may decay after a very short time, this does not imply that a superposition of different space-time geometries should also decay - and, again, sometimes an unstable state does not decay at all under special circumstances. To sum up, Penrose's argument by analogy only has a very limited force, and it is not strong enough to establish a necessary connection between wavefunction collapse and the conflict between quantum mechanics and general relativity.

Thirdly, it can be further argued that the conflict between quantum mechanics and general relativity does not necessarily lead to wavefunction collapse. The key is to realize that the conflict also needs to be resolved before the wavefunction collapse finishes, and when the conflict has been resolved, the wavefunction collapse will lose its basis relating to the conflict. As argued by Penrose, a quantum superposition of different space-time geometries and its evolution are both ill-defined due to the fundamental conflict between the general covariance principle of general relativity and the superposition principle of quantum mechanics. The ill-definedness seems to require that the superposition must collapse into one of the definite space-time geometries, which has no problem of ill-definedness. However, the wavefunction collapse seems too late to save the superposition from the "suffering" of the ill-definedness during the collapse. In the final analysis, the conflict or the problem of ill-definedness needs to be solved *before* defining a quantum superposition of different space-time geometries and its evolution. In particular, the possible collapse evolution of the superposition also needs to be consistently defined, which again indicates that the wavefunction collapse does not solve the problem of ill-definedness. On the other hand, once the problem of ill-definedness is solved and a consistent description obtained (however this is still an unsolved issue in quantum gravity),

the wavefunction collapse will lose its connection with the problem[2]. Therefore, contrary to Penrose's expectation, it seems that the conflict between quantum mechanics and general relativity does not entail the existence of wavefunction collapse.

Even though Penrose's gravity-induced collapse argument may be problematic, it is still possible that the wavefunction collapse is a real physical process (Gao 2011). Therefore, Penrose's suggestions for the collapse time formula and collapse states also need to be examined as some aspects of a phenomenological model. To begin with, let's analyze Penrose's collapse time formula Eq. (1), according to which the collapse time of a superposition of two mass distributions is inversely proportional to the gravitational self-energy of the difference between the two mass distributions. As we have argued above, the analogy between such a superposition and an unstable state in quantum mechanics does not exist, and gravity does not necessarily induce wavefunction collapse either. Thus this collapse time formula, which is based on a similar application of Heisenberg's uncertainty principle to unstable states, will lose its original physical basis. In particular, the appearance of the gravitational self-energy term in the formula is in want of a reasonable explanation (see below). In fact, it has already been shown that this gravitational self-energy term does not represent the ill-definedness of time-translation operator in the strictly Newtonian regime (Christian 2001). In this regime, the time-translation operator can be well defined, but the gravitational self-energy term is not zero. Besides, as Diósi (2007) pointed out, the microscopic formulation of Penrose's collapse time formula also meets the cut-off difficulty.

Next, let's examine Penrose's suggestion for the collapse states. According to Penrose (1998), the collapse states are the stationary solutions of the Schrödinger-Newton equation:

$$i\hbar\frac{\partial\psi(\mathbf{x},t)}{\partial t}=-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{x},t)-Gm^2\int\frac{|\psi(\mathbf{x}',t)|^2}{|\mathbf{x}-\mathbf{x}'|}d^3\mathbf{x}'\psi(\mathbf{x},t)+V\psi(\mathbf{x},t), \quad (2)$$

where m is the mass of a quantum system, V is an external potential, G is Newton's gravitational constant. The equation describes the gravitational self-interaction of a single quantum system, in which the mass density $m|\psi(x,t)|^2$ is the source of the classical gravitational potential. As we have argued in a previous paper (Gao 2011), although a quantum system has a mass distribution that is measurable by protective measurement, the distribution is not real but effective; it is formed by the ergodic motion of a localized particle with the total mass of the system. Therefore, there does not exist a gravitational self-interaction of the mass distribution. This conclusion can also be reached by another somewhat different argument. Since charge always accompanies mass for a charged

[2]Note that if the problem of ill-definedness cannot be solved in principle for the superpositions of very different space-time geometries, then the wavefunction collapse may be relevant here. Concretely speaking, if the superpositions of very different space-time geometries cannot be consistently defined even in principle, then these superpositions cannot exist and they must have collapsed into one of the definite space-time geometries before being formed from the superpositions of minutely different space-time geometries. In this case, the large difference of the space-time geometries in the superposition will set an upper limit for wavefunction collapse. Though the limit may be loose, it does imply the existence of wavefunction collapse. However, this possibility may be very small, as it seems that there is always some kind of approximate sense in which two different spacetimes can be pointwise identified.

particle such as an electron[3], the existence of the gravitational self-interaction, though which is too weak to be excluded by present experiments (Salzman and Carlip 2006), entails the existence of a remarkable electrostatic self-interaction of the particle, which already contradicts experiments (Gao 2011). This analysis poses a serious objection to the Schrödinger-Newton equation and Penrose's suggestion for the collapse states[4].

Lastly, we briefly discuss another two potential problems of Penrose's collapse scheme. The first one is the origin of the randomness of collapse results. Penrose did not consider this issue. If the collapse is indeed spontaneous as implied by his gravity-induced collapse argument, then the randomness cannot result from any external influences such as an external noise field, and it can only come from the studied quantum system and its wave function (Gao 2011). The second problem is energy non-conservation. Although Penrose did not give a concrete model of wavefunction collapse, he thought that the energy uncertainty E_Δ may cover such a potential non-conservation, leading to no actual violation of energy conservation (Penrose 2004). However, Diósi (2007) pointed out that the von-Neumann-Newton equation, which may be regarded as one realization of Penrose's collapse scheme, does not conserve the energy. If the principle of conservation of energy is indeed universal as widely thought, then the spontaneous collapse models that violate energy conservation will have been excluded[5].

To sum up, we have argued that Penrose's argument for gravity's role in wavefunction collapse is debatable. However, it is still possible that the wavefunction collapse is a real physical process, though its origin remains a deep mystery.

References

[1] Christian, J. (2001). Why the quantum must yield to gravity. In: Physics Meets Philosophy at the Planck Scale, C. Callender and N. Huggett (ed.). Cambridge: Cambridge University Press. p. 305.

[2] DeWitt, C. and Rickles, D. (ed.) (2011). The Role of Gravitation in Physics: Report from the 1957 Chapel Hill Conference. Max Planck Research Library for the History and Development of Knowledge, Vol. 5.

[3] Diósi, L. (1989). Models for universal reduction of macroscopic quantum fluctuations. Phys. Rev. A 40, 1165-1173.

[4] Diósi, L. (2007). Notes on certain Newton gravity mechanisms of wave function localisation and decoherence. J. Phys. A: Math. Gen. 40, 2989-2995.

[3]However, the concomitance of mass and charge in space for a charged particle does not necessarily require that they must satisfy the same law of interaction. For example, the fact that electromagnetic fields are quantized in nature does not necessarily imply that gravitational fields must be also quantized.

[4]Since the Schrödinger-Newton equation is the non-relativistic realization of the typical model of semiclassical gravity, in which the source term in the classical Einstein equation is taken as the expectation of the energy momentum operator in the quantum state, this analysis also poses a serious objection to the approach of semiclassical gravity.

[5]It has been demonstrated by a concrete model that wavefunction collapse may conserve energy (Gao 2011, ch.4).

[5] Feynman, R. (1995). Feynman Lectures on Gravitation. B. Hatfield (ed.), Reading, Massachusetts: Addison-Wesley.

[6] Gao, S. (2010). On Diósi-Penrose criterion of gravity-induced quantum collapse. Int. J. Theor. Phys. 49, 849-853.

[7] Gao, S. (2011). Interpreting Quantum Mechanics in Terms of Random Discontinuous Motion of Particles. http://philsci-archive.pitt.edu/8987.

[8] Penrose, R. (1996). On gravity's role in quantum state reduction. Gen. Rel. Grav. 28, 581.

[9] Penrose, R. (1998). Quantum computation, entanglement and state reduction. Phil. Trans. R. Soc. Lond. A 356, 1927.

[10] Penrose, R. (2004). The Road to Reality: A Complete Guide to the Laws of the Universe. London: Jonathan Cape.

[11] Rovelli, C. (2011). "Forget time": Essay written for the FQXi contest on the Nature of Time. Found. Phys. 41, 1475-1490.

[12] Salzman, P. J. and Carlip, S. (2006). A possible experimental test of quantized gravity. arXiv: gr-qc/0606120.

[13] Yablonovitch, E. (1987). Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett. 58, 2059.