A Final Summary

Quantum fluctuations, particles and entanglement: solving the quantum measurement problems

Quantum measurement problems

Quantum mechanics is described by the following few formulas,

$$[x_i, p_j] = i\hbar\, \delta_{ij}\,, \qquad \Delta x \Delta p \geq \frac{\hbar}{2}\,;$$

$$i\hbar\frac{d\psi}{dt} = H\psi\,;$$

$$H\psi_n = E_n\psi_n\,, \qquad H = \frac{\mathbf{p}^2}{2m} + V\,, \qquad \mathbf{p} = -i\hbar\nabla\,. \tag{1.1}$$

The rest is, essentially, just details.

In spite of the fundamental remarkable simplicities, quantum mechanics (QM) has led to

An amazing success and detailed confirmation in atomic physics, ushering us into a century of impressive advance in basic scientific knowledge and modern technologies, unprecedented in human history. We still live in the wake of this ongoing revolution.

Further pursuit into subatomic physics has eventually led to the standard model of the fundamental interactions, a non-Abelian gauge theory based describing very accurately the strong (nuclear) interactions and electroweak interactions, down to the distance scales of order of $O(10^{-18})$ cm.

One cannot forget about the beautiful phenomena in condensed matter physics, such as superconductivity, quantum Hall effect, Bose-Einstein condensation of cold atoms, etc. etc.

We can certainly talk about

100 years of extraordinary scientific achievements.

But, at the same time, the QM predictions are, supposedly, given by the probabilistic (Born's) rule.

And this has led to

100 years of uneasy feeling and doubts,

that something fundamental is missing in our understanding of QM.

These are often expressed as various (apparent) puzzles or conceptual difficulties, which are all loosely called the "quantum measurement problems".

They are actually a collection of different questions:

(a) Wave function collapse?

(b) Macroscopic superposition of states?

(c) Forever branching manyworlds tree?

(d) Spontaneous collapse of micro state, of the experimental device, and of the world?

(e) Wave function just a bookkeeping device?

(f) Quantum jumps vs Schrödinger equation?

(g) Born's rule from the Schrödinger equation involving everything?

(h) Quantum nonlocality, Hidden variables?

(i) How is it possible that the fundamental, complete theory of physics is incapable of unique prediction?

The following summary discussion will address all of them.

The actual measurement devices can vastly differ in their nature, size, materials used, and technologies employed, from a simple photographic plate, cloud and bubble chambers filled with liquid or vapor, spark chambers and MWPC made with metal plates, wires and gas, the neutrino detectors made of a huge tank of pure water and tens of thousands of photomultipliers, to the state-of-the-art silicon detectors and some future apparatus which uses superconducting materials for dark matter search.

Such an enormous diversity of the experimental devices requires that an equally vast simplification be made, to capture the essence of quantum measurements.

Indeed, independently of the details, a good experimental device faithfully reflects the quantum fluctuations of the system, described by the wave function $|\psi\rangle$.

Indeed, the result of each single experiment is, in general, apparently random and unpredictable.

We remember, of course, that an exception occurs when the state ψ is one of the eigenstates of the quantity F, with $F = f_m$, i.e., $|\psi\rangle = |m\rangle$.

In such a case, a good experiment produces the same result $F = f_m$ every time.

Yet, the information encoded in the wave function $|\psi\rangle$ manifests itself as the expectation values for any variable F (and all functions thereof),

$$\bar{F} \equiv \langle \psi | F | \psi \rangle \tag{1.2}$$

that is, as a (frequency-) average of the experimental outcomes, $\bar{f}_n$,

$$\bar{f}_n = \sum f_n \, \mathcal{P}_n , \qquad \mathcal{P}_n = |\langle n | \psi \rangle|^2 \tag{1.3}$$

where $F |n\rangle = f_n |n\rangle$.

$\mathcal{P}_n$ is the relative frequency for f_n.

Such a prediction of quantum mechanics is verified by countless experiments.

That is, "QM works".

We note that, throughout, we use simply the word "variable", for a dynamical variable, physical quantity, such as energy, momentum, angular momentum, position, etc., which is measurable (known as an observable).

Also, we use the same symbol F for the physical variable itself and for the self-adjoint operator representing it.

You should not have any difficulty telling which is meant, each time.

The prediction (1.3) follows, if one assumes that the probability of each single measurement of F in the state ψ to give f_n, is given by $P_n = |\langle n|\psi\rangle|^2$, that is, by Born's rule.

Born's rule is always presented as one of the fundamental postulates of quantum mechanics

So what is the problem?

Well, "the quantum measurement problems", (a) ~ (i), above!

Throughout, we will denote the (normalized) relative frequency with the calligraphy font $\mathcal{P}_n$, and write the standard probability by using the normal font P_n.

Although they express mathematically the same numbers, they differ conceptually, and lead to distinct logical conclusions.

Three aspects of QM

Before analyzing the measurement processes themselves later, we must first discuss briefly three familiar aspects of QM.

Though singly well known, clarifying their precise roles in the measurement processes will be essential in the following discussions.

Expectations values versus Born's rule

To fix the idea, we consider the system described by the wave function,

$$|\psi\rangle = \sum_n c_n |n\rangle , \qquad F|n\rangle = f_n|n\rangle , \qquad \sum_n |c_n|^2 = 1 , \qquad (2.1)$$

where F is the variable we are interested in.

Repeated, identical measurement of F will give an average, (1.3).

Now, (1.3) certainly follows from Born's rule.

In other words, the latter is a sufficient condition for the former.

The problem is: is Born's rule also necessary?

$$O(\bar{f}_n)$$

The answer is yes, if the following equality holds

$$O(\bar{f}_n) \equiv \sum_n \mathcal{P}_n O(f_n) = \langle\psi|O(F)|\psi\rangle \equiv \sum_n |c_n|^2 O(f_n) , \qquad (2.2)$$

for a general function O(F) of F.

Then it can be shown that

$$\mathcal{P}_n = |c_n|^2 = P_n \,.$$ (2.3)

Pure and mixed states

A system described by a wave function (or a state vector) ψ is a pure state.

It contains the complete knowledge of the system.

Otherwise, the system is in a mixed state (or a mixture).

It is described by the density matrix, ϱ.

For instance, a subsystem $A \subset \Sigma$ of a closed system Σ, described by the wave function,

$$|\Psi\rangle = \sum_{n,\alpha} c_{n,\alpha} |n\rangle |\alpha\rangle \,, \qquad \sum_{n,\alpha} |c_{n,\alpha}|^2 = 1$$ (2.4)

the expectation value of G (pertinent to the subsystem A) is given by

$$\bar{G} = \langle \Psi | G | \Psi \rangle = \mathrm{Tr}\, \mathbb{G}\rho \,, \qquad \mathbb{G}_{mn} \equiv \langle m | G | n \rangle \,,$$ (2.5)

where the density matrix is given in this case by

$$\rho_{nm} \equiv \sum_\alpha c_{n,\alpha} c_{m,\alpha}^* \,.$$ (2.6)

More generally the density matrix represents any sort of ignorance about the system, and will have forms different from (2.6), but the trace formula

$$\bar{G} = \mathrm{Tr}\, \mathbb{G}\rho \,,$$ (2.7)

is valid always.

Factorization vis-a-vis entanglement

The question is, symbolically, this: how can we study a single hydrogen atom, an electron, etc., when we know that the wave function involving identical fermions must be antisymmetrized (Fermi-Dirac (FD) statistics)?

The answer is the lack of the spatial support (effect on a calculation) in the wrong component.

For instance, for an electron in the laboratory and another in the Sun, the wave function will look like

$$\Psi = \frac{1}{\sqrt{2}}\left(\psi_{lab}(\mathbf{r_1})\psi_{sun}(\mathbf{r_2}) - \psi_{lab}(\mathbf{r_2})\psi_{sun}(\mathbf{r_1})\right) \sim \psi_{lab}(\mathbf{r_1}), \tag{2.8}$$

for

$$\mathbf{r_1} \in Lab. , \qquad \mathbf{r_2} \in Sun \tag{2.9}$$

as clearly there is no support for the wrong component, $\psi_{lab}(\mathbf{r_2})$ or $\psi_{sun}(\mathbf{r_1})$.

One sees how a pure state emerges as the result of factorization.

If one takes two hydrogen atoms (two electrons) ~ 2m apart, in the next room in the laboratory, instead of the second in the Sun, the factorization failure (the effect of the wrong component) is still a very small effect,

$$\sim 10^{-18}, \tag{2.10}$$

We note that any systems which have interacted in the past are entangled, and in general, in a mixed state.

Factorization and entanglement both characterize quantum mechanics universally.

Note in particular that

Factorization *works remarkably strongly; it makes QM a workable, sensible, physics theory, in spite of the fact that everything is in principle entangled in our universe. It is factorization that makes the pure state a significant concept, both from theoretical and experimental points of view;*

Entanglement *leads to the fascinating, characteristic phenomena in QM, such as quantum nonlocality, violation of Bell's and CHSH inequalities, quantum cryptgraphy, etc. At the same time, the uncontrolled or uncontrollable entanglement with the unobserved systems, is responsible for decoherence, and the emergence of classical physics.*

A reflection

In principle, we always live in a mixed state: everything is interrelated, and interacted or interacting with each other, in our universe.

It is the notion of a pure quantum state, described by a wave function ψ, which is extraordinary, and exceptional.

They arise as a result of factorization.

These pure states are carefully prepared by experimentalists in many small bubbles in the world, called physics laboratories, equipped with a good clean room and sophisticated (and expensive) vacuum and cryogenic technologies.

But they occur also naturally, on the earth, in the form of radioactivity, which produces α (α particles), β (electrons) or γ (photons) rays.

They also fall from the sky, in the form of cosmic rays (energetic protons, pions, etc.).

Then there was light.

Light, as it turns out, is a gas of non-interacting, thus, free, photons.

They are all pure quantum states.

In a hindsight, we see why these phenomena (providing pure quantum states for free!) have led to the discovery of quantum physics by Planck, at the dawn of the 20th century.

The measurement processes

Now we come to our main problem: the quantum measurement processes.

Effective spacetime localization - state-vector reduction

The measurement of the variable F in the state

$$|\psi\rangle = \sum_n c_n |n\rangle , \qquad F|n\rangle = f_n|n\rangle , \tag{3.1}$$

is often assumed to proceed as

$$|\psi\rangle \otimes |\Phi_0\rangle \otimes |Env_0\rangle \quad = \quad \left(\sum_n c_n|n\rangle\right) \otimes |\Phi_0\rangle \otimes |Env_0\rangle \tag{3.2}$$

$$\longrightarrow \left(\sum_n c_n|n\rangle \otimes |\Phi_n\rangle\right) \otimes |Env_0\rangle \tag{3.3}$$

$$\longrightarrow \sum_n c_n|n\rangle \otimes |\Phi_n\rangle \otimes |Env_n\rangle , \tag{3.4}$$

where in the first step (3.3) the experimental device reads the measurement results, f_n, and in the final step (3.4) the "environment" comes be aware of that result (the experimentalist has seen the result on her/his computer screen).

These formulas appear to suggest a coherent superposition of distinct macroscopic states leading to paradoxes, debates and endless confusion.

Actually, the factorized form in (3.2)-(3.4), with various $\otimes$ symbols, is not correct, except for the factorization of $|\psi\rangle$ in the first line, namely, before the experiment.

We note that the factorized form $|\Phi\rangle \otimes |\text{Env}\rangle$ is incorrect, even before the measurement.

The air molecules, the casing of the device, etc., all show that $|\Phi\rangle$ and $|\text{Env}\rangle$ are actually entangled.

Not only, the boundaries between $|\Phi\rangle$ and $|\text{Env}\rangle$ are badly defined, and to be regarded as just a convention.

Nevertheless, any experimentalist knows what the relevant part of the device is.

She (or he) knows perfectly well that there is no need to push the boundary between the measuring device $|\Phi\rangle$ and the rest of the world $|\text{Env}\rangle$ up to inside the human brain, in order to ensure good, precise measurement results.

Keeping this state of matter in mind, we denote the measurement-device-environment entangled "state", as

$$|\Phi; Env\rangle , \tag{3.5}$$

below.

The measurement process can be represented more appropriately as

$$|\psi\rangle \otimes |\Phi_0; Env_0\rangle \quad = \quad \left(\sum_n c_n |n\rangle\right) \otimes |\Phi_0; Env_0\rangle \tag{3.6}$$

$$\longrightarrow \quad \sum_n c_n |\tilde{\Phi}_n; Env_0\rangle \tag{3.7}$$

$$\longrightarrow \quad \sum_n c_n |\tilde{\Phi}_n; Env_n\rangle . \tag{3.8}$$

where $\bar{\Phi}_n$ stands for the entangled state of the microsystem-apparatus with the reading of the measurement result, f_n.

The exact timing of passage from the first stage of measurement-registering of the result on the apparatus (3.7) to the second (3.8) (e.g., the moment in which the experimentalist sees the result on her or his computer screen; others read about it in Physical Review), is entirely immaterial.

We note that the device-enviroment "state", $|\Phi_0; Env_0\rangle$, can never be identical, at two different measurement instants as we will discuss later.

Emergence of the mixed state and the state-vector reduction (the "wave function collapse") occurs as follows.

Just before the measurement, the system is described by a wave function of factorized form by assumption,

$$|\Psi\rangle_{t=0_-} = \left(\sum_n c_n |n\rangle\right) \otimes |\Phi_0; Env_0\rangle . \tag{3.9}$$

The expectation value of any generic variable G in this state (prior to the measurement) is given by

$$\bar{G} = \langle \Psi | G | \Psi \rangle = \text{Tr}\,\mathbb{G}\rho\,, \qquad (\mathbb{G})_{mn} \equiv \langle m | G | n \rangle\,, \qquad (3.10)$$

with the density matrix having the form

$$\rho = \begin{pmatrix} |c_1|^2 & c_1 c_2* & c_1 c_3^* & \cdots & \cdots \\ c_2 c_1^* & |c_2|^2 & c_2 c_3^* & \cdots & \\ c_3 c_1^* & & & \ddots & \\ \vdots & & & & |c_n|^2 \\ & & & & & \ddots \end{pmatrix}, \qquad (3.11)$$

$$\mathbb{F}$$

characteristic of a pure state, i.e., $\varrho_{nm} = c_n c_m^*$.

The normalization condition

$$\langle \Phi_0; Env_0 | \Phi_0; Env_0 \rangle = 1 \qquad (3.12)$$

has been used. In the case of the particular variable F (whose eigenstates are |n⟩'s), $\mathbb{F}$ is diagonal, and its quantum average is given by the known formula

$$\bar{F} = \text{Tr}(\mathbb{F}\rho) = \sum_n f_n |c_n|^2\,. \qquad (3.13)$$

As soon as the microsystem gets into contact with the experimental device, and the measurement events (a chain ionization process, hadronic cascade, etc.) have taken place, the total wave function takes an entangled form, (3.7)

$$|\Psi\rangle_{t=0_+} = \sum_n c_n |\tilde{\Phi}_n; Env_0\rangle .$$

(3.14)

$\bar{\Phi}_n$ denotes the entangled microsystem-apparatus state with the reading of the measurement result, f_n.

Now, the wave function describing $\bar{\Phi}_m$ (the aftermath of a measurement, with the result, $F = f_m$) and that for $\bar{\Phi}_n$ (the aftermath of a measurement, with the result, $F = f_n$), corresponding to two distinct spacetime events, have no overlapping spacetime support, as illustrated below.

(3.14) is a mixed state.
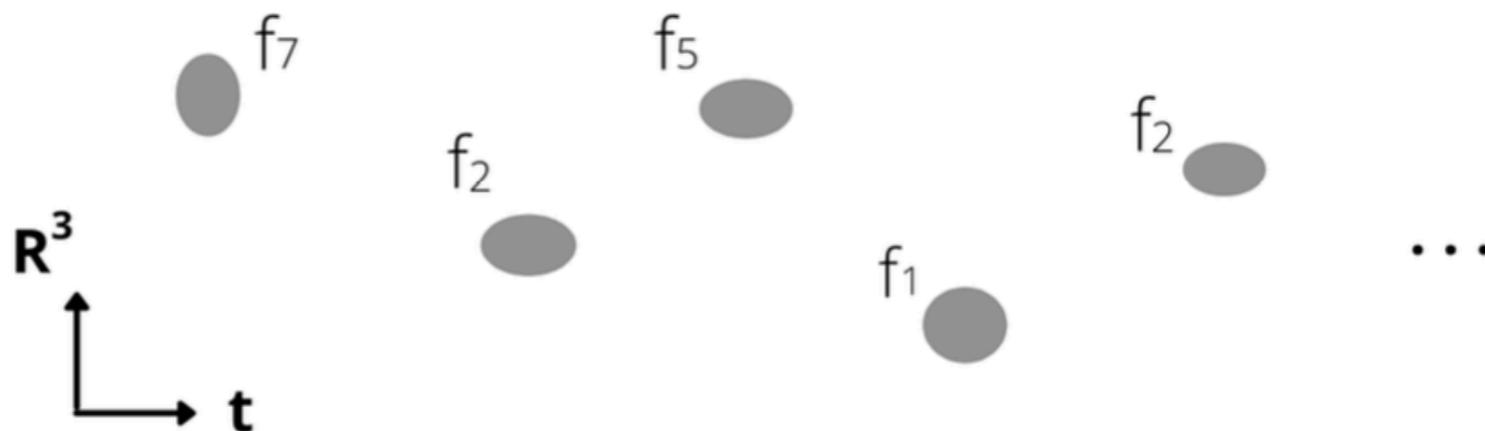
Measurement of a variable F. Each blob describes a single measurement event (a chain ionization reaction, hadronic cascade, etc.) occurring at a localized region in spacetime, with the experimental results, $F = f_7, f_2, f_5, f_1, f_2$, etc. The wave functions describing the different measurement events $\bar{\Phi}$ have no overlapping spacetime supports, therefore are orthogonal.

Treating the formula (3.14) as a mixed state, and not a coherent superposition of states, eliminates at once the contradiction found in many papers.

The state after the measurement, (3.14), cannot be a coherent superposition.

The expectation values of a variable G are then given by

$$\bar{G} = \langle \Psi | G | \Psi \rangle = \sum_{m,n} c_m^* c_n \langle \tilde{\Phi}_m; Env_0 | G | \tilde{\Phi}_n; Env_0 \rangle \; ; \tag{3.15}$$

but the fact $\bar{\tilde{\Phi}}_m$ and $\bar{\tilde{\Phi}}_n$ (m ≠ n) have no common spacetime support, means that for any local operator G, orthogonality relations,

$$\langle \tilde{\Phi}_m; Env_0 | G | \tilde{\Phi}_n; Env_0 \rangle = 0 \; , \qquad m \neq n \; , \tag{3.16}$$

hold.

Consequently, $\bar{G}$ is given by the sum of the diagonal terms

$$\bar{G} = \sum_n |c_n|^2 G_{nn} \; , \qquad G_{nn} = \langle \tilde{\Phi}_n; Env_0 | G | \tilde{\Phi}_n; Env_0 \rangle \; . \tag{3.17}$$

This means that the density matrix has been reduced to a diagonal form,

$$\rho = \begin{pmatrix} |c_1|^2 & & & & & \\ & |c_2|^2 & & & & \\ & & \ddots & & & \\ & & & |c_n|^2 & & \\ & & & & \ddots & \end{pmatrix} . \tag{3.18}$$

For the particular case of the variable F , we recover the standard prediction,

$$\bar{F} = \sum_n |c_n|^2 f_n , \qquad \therefore \quad \mathcal{P}_n = |c_n|^2 , \qquad (3.19)$$

where $\mathcal{P}_n$ are the normalized relative frequencies for different outcomes f_n.

The emergence of the mixed state (3.18) is often attributed to the fact that a typical experimental device is made of a macroscopic (difficult-to-specify) number of atoms and molecules, and it is not possible to keep track of the phase relations among different terms in (3.7) to any significant extent.

Also, the macroscopic device Φ, evolving in t entangled with Env, can never be in an identical quantum state $|\Phi_0; Env_0\rangle$ at two different measurement instants.

This is in stark contrast with the identical quantum state for the microsystem, $|\psi\rangle$, which can be and is indeed produced via e.g., a repeatable experiment (state preparation - see earlier) for each measurement.

The observations above are both certainly correct, but we need another, crucial ingredient for the decoherence in the measurement processes: the lack of the common spacetime supports in the wave functions, and the consequent orthogonality and decoherence among terms corresponding to different measurement results, (3.16).

Importance of this is that it *implies that the result of each measurement event is a state-vector reduction*,

$$\left(\sum_n c_n |n\rangle \right) \otimes |\Phi_0; Env\rangle \Longrightarrow |\tilde{\Phi}_m; Env\rangle : \qquad (3.20)$$

i.e., with a single term present, the instant after the measurement (e.g., with F = f_m).

This state of affair is perceived by us as a *"wave-function collapse"*.

Summarizing:

A: The spacetime eventlike nature of the triggering particle-measuring-device interactions introduces an effective spacetime localization of each measurement event;

B: At the moment the microsystem - a pure quantum state $|\psi\rangle$ - gets into contact with the measurement device-environment state $|\Phi_0; Env_0\rangle$ which is a mixed state, factorization of $|\psi\rangle$ gets lost rapidly, and an entangled, mixed (classical) state $|\Phi_n; Env_0\rangle$ with the unique recording, $F = f_n$, is generated.

The result is the state-vector reduction, (3.20).

The effective spacetime localization and apparent wave-function collapse, can be nicely visualized in the "double-slit" experiments by Tonomura et. al, see figure.

The pictures correspond, from the top to the bottom, to successive exposures to 10, 100, 3000, 20000 and 70000 electrons, respectively.
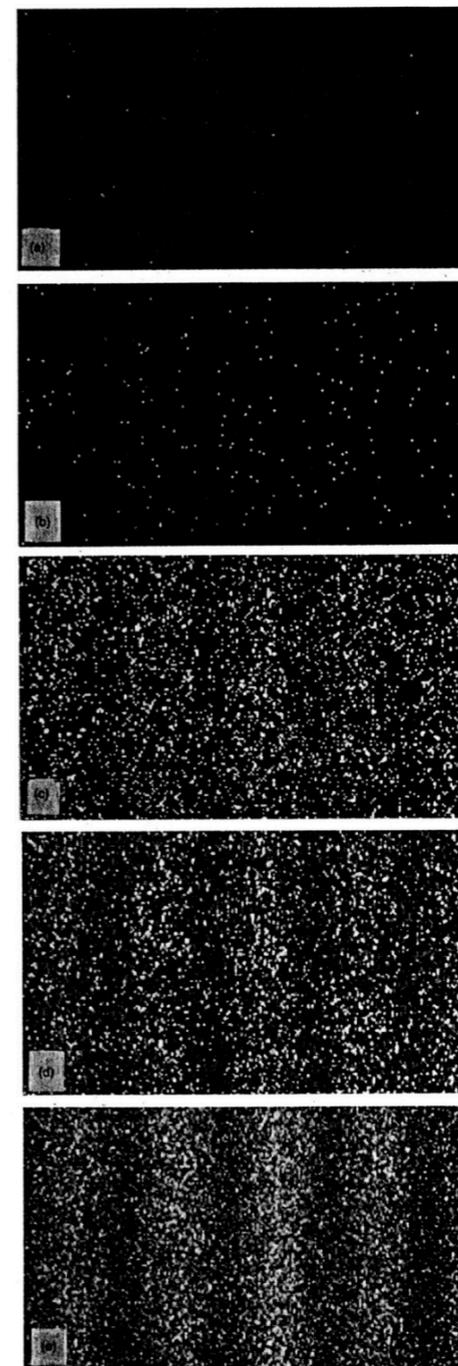
Repeatable, nonrepeatable and intermediate types of state reductions

There are actually several types of measurements.

The first is known as repeatable, or of the first kind: it proceeds as

$$\left(\sum_n c_n|n\rangle\right) \otimes |\Phi_0; Env_0\rangle \longrightarrow |m\rangle \otimes |\Phi_m; Env\rangle \,, \tag{3.21}$$

which corresponds to the expression often found in a QM textbook,

$$|\psi\rangle = \left(\sum_n c_n|n\rangle\right) \longrightarrow |m\rangle \,. \tag{3.22}$$

In these experiments, the measuring device has acquired and recorded the information about the microsystem under observation, but the latter remains unscathed, as a factorized, pure state.

These do represent special class of measurements, but they are actually neither rare nor particularly difficult to realize.

A variant of the Stern-Gerlach experiment (see figure) is a good example.

Stern-Gerlach experiment
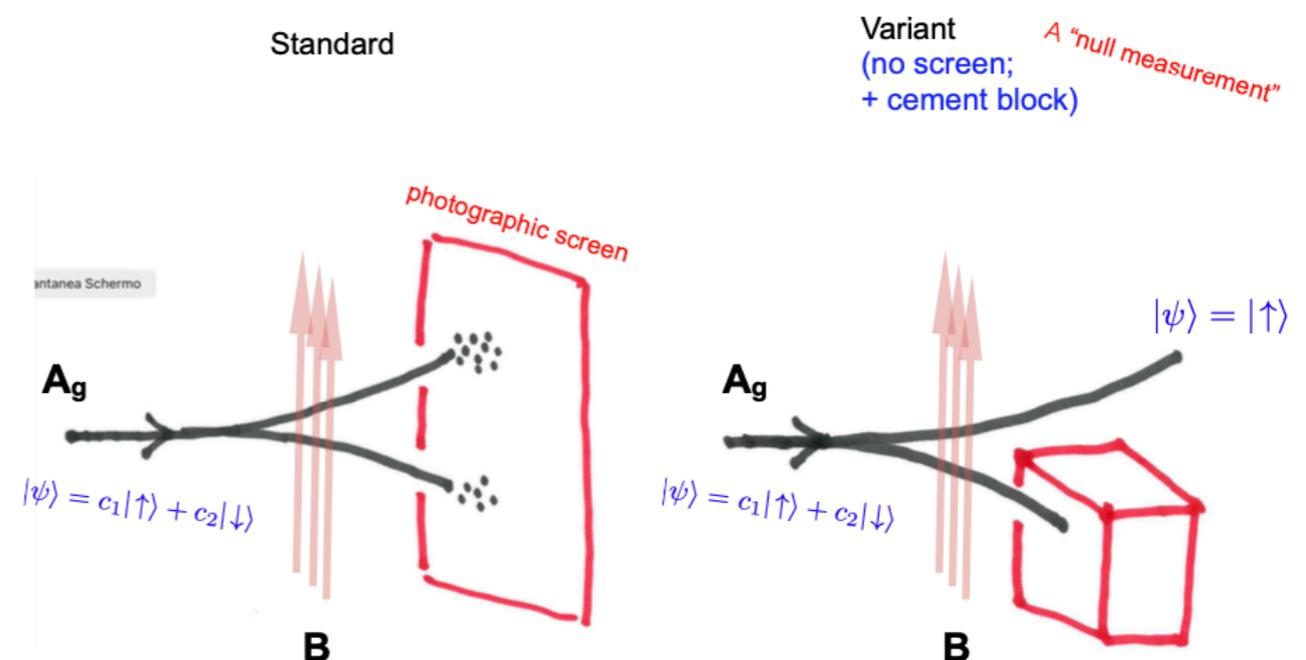
A variant (right) of the standard SG experiment (left). A cement block stops the lower beam containing the spin-down atoms. No measurement is done (known as a "null measurement"), but the atoms coming out of the region consist purely of spin-up state, which can be used as the initial wave function $|\psi(0)\rangle = |\uparrow\rangle$.

The very possibility of the repeatable measurements is actually of fundamental importance for QM.

They allow the preparation of a desired, identical, pure quantum state $|\psi\rangle$ as the initial state, whenever we wish.

This is important because in QM the experimental check of the theory is done necessarily with many repeated, identical experiments, as QM predicts only the relative frequencies for various outcomes.

Another, related fact is that the atoms of the same kind, in their ground state, are all rigorously identical, in contrast to classical particles.

Such a characteristics of quantum mechanical world is at the basis of the regularity of the macroscopic world, such as crystalline structures and extremely precise biological phenomena such as the reproduction.

In contrast, a macroscopic body made of, say, 10^{25} atoms and molecules (e.g., an experimental device), can never be in an identical quantum state at different times or in different laboratories, even though they might look identical at macroscopic scales.

This is in fact one of the underlying reasons for the state-vector reduction (see the discussion above, (3.20)).

The second type of the measurement may be simply termed non repeatable.

This is a general type of the measurement - state reduction, as in (3.20).

The result $F = f_m$ is recorded, but the original microsystem itself (e.g., the electron in Tonomura's experiment) simply gets lost, in $\left|\bar{\Phi}_m; Env\right\rangle$.

There are also third, intermediate types of measurements.

The detection of particle tracks in Wilson, or bubble chambers, or in the silicon vertex detectors, is an example of the measurement of this sort.

Note that in these experiments the microscopic particle interacts with the atoms in the devise, generates a local chain ionization cloud, but its information (the position or momentum) does not get lost completely, creating successive ionization clouds (particle tracks) on its way.

This kind of processes is particularly important in the momentum or energy measurement.

Unitarity and linearity

The microscopic state ψ

$$|\psi\rangle = \sum_n c_n |n\rangle \, , \tag{3.23}$$

if *left undisturbed*, evolves in time according to the Schrodinger equation,

$$i\hbar \frac{d}{dt} |\psi\rangle = H |\psi\rangle \, , \tag{3.24}$$

(H being the Hamiltonian) or as

$$|\psi\rangle \longrightarrow U |\psi\rangle = \sum_n c_n U |n\rangle \, , \qquad U = e^{-iHt/\hbar} \, , \qquad U^\dagger U = 1 : \tag{3.25}$$

it is a linear, unitary evolution.

Unitarity means that

$$\langle \psi | \psi \rangle = \sum_n |c_n|^2 = 1 \,, \tag{3.26}$$

i.e., the state norms are conserved in time.

Linearity means (3.25), i.e., superposition of different states in (3.23) continue to be coherent superposition of the corresponding states.

During the measurement process, formally written as (3.7), the time evolution might still look unitary and linear.

However, as we have seen, the wave functions associated with different terms in (3.7) do not overlap in spacetime, so it represents actually a mixed state.

The coherent superposition of states is no longer there, see (3.16).

Most significantly, the real time evolution of the system is the state-vector reduction, (3.20), meaning that *linearity in the sense of (3.25) is lost during the measurement*.

On the other hand, unitarity is maintained: the sum of all possible outcomes, occurring at different measurements at different times, adds up to unity of the total normalized frequencies, corresponding to the "norm" of the state, (3.7).

In terms of the relative frequencies for different outcomes, $\mathcal{P}_n$, unitarity means

$$\sum_n \mathcal{P}_n = 1 \,, \qquad \sum_n |c_n|^2 = 1 \,, \tag{3.27}$$

even though different (sometimes the same, but repeated) results f_n refer to distinct measurements made at different times.

Note that this is indeed how experimentalists view the meaning of unitarity.

Unitarity means that $\mathcal{N}_n/\mathcal{N}$, to be identified with the theoretical formula, $\mathcal{P}_n = |c_n|^2$, should satisfy

$$\sum_n \mathcal{N}_n/\mathcal{N} = 1 , \tag{3.28}$$

which might look trivial.

Note that $\mathcal{N}_n$ is the number of times the experimentalist finds the result $F = f_n$, $n = 1, 2, \ldots$; $\mathcal{N}$ is the total number of the measurements made.

However, this contains an implicit, important assumption that the experimental device has no systematic bias, and registers all possible results f_n with equal efficiency and with no losses.

Only in such an ideal measurement setting we can expect that the experimental results $\mathcal{N}_n/\mathcal{N}$ will approach the theoretical prediction, $\mathcal{P}_n$, in the limit of large $\mathcal{N}$ (e.g., Tonomura's experiment).

Note that linearity and unitarity are two distinct concepts in quantum mechanics (e.g., the momentum operator is linear but not unitary).

Dirac noted that the evolution operator must satisfy both, which are two independent requirements.

They are both automatically met, once the choice $U = e^{-iHt/\hbar}$ is made with a self-adjoint Hamiltonian operator, and as long as the pure-state evolution (3.25) before the measurement, is concerned.

It might be of some interest to compare the concept of unitarity in the measurement processes as described here, with a more abstract one, in the conventional thinking based on Born's rule.

The latter means in fact

$$\sum_n P_n = 1 \,, \tag{3.29}$$

i.e., that the total probability for a single experiment to give all possible results f_n, is unity, and that it is conserved in time.

The request (3.29), from the mathematical, logical point of view, appears quite indispensable, and indeed has always been considered as a sacrosanct principle in the traditional approach to quantum mechanics.

However, once P_n is translated into the physically directly meaningful quantity, $\mathcal{P}_n$ (= P_n), the expected relative frequencies for various results f_n which will be found in distinct measurements and at different times, unitarity as an absolute sacred principle may appear to lose part of the aura surrounding it.

To conclude, one must be cautious in applying the concepts such as linearity or unitarity of the time evolution, valid in the context of (isolated) pure states, to the measurement processes where entanglement between the microsystem and the measurement device as well as with the whole world, plays an essential role.

Their consequences in these, complicated mixed states may well (and indeed, do) look differently from what one is accustomed to, from the study of pure quantum states kept in isolation, (3.24), (3.25).

Measurements, emergence of classical physics, macroscopic quantum states, and models of wave-function collapses

The state after a measurement, $\left|\bar{\Phi}_m; Env\right\rangle$ (in (3.20), is an entangled, classical state, with the recording of the measurement result.

The basic reason for that is the so-called enviornment-induced decoherence.

Emergence of classical physics in this context, is certainly one of the important ingredients in the understanding of the quantum measurement processes (the unique aftermath of each measurement - the wave function collapse).

However, it is useful to consider the whole problem from a wider perspective of what may be called "great twin puzzles of physics today": the general quantum measurement problems on the one hand, and emergence of the classical mechanics from QM, on the other.

Even though these two classes of the problems are often discussed together, they are actually mostly independent issues.

It is better to discuss them separately, and independently.

Here we are concerned with the first of them, the general measurement problems; for the second (how Newton's law emerges from QM), there are many papers.

For instance, we know that sufficiently close to T = 0, any matter, in any state, is quantum mechanical.

At T = 0 any system is in the unique ground state (Planck-Nernst law, or the first law of thermodynamics).

At T > 0 but sufficiently close to the absolute zero temperature, a macroscopic body can be in a coherent superposition,

$$|\Psi\rangle = c_1|\Psi_1\rangle + c_2|\Psi_2\rangle \tag{3.30}$$

where $|\Psi_1\rangle$ and $|\Psi_2\rangle$ are macroscopically distinct states.

There have been many experimental efforts to realize macroscopic (or mesoscopic) quantum states at very low temperatures.

To talk about the macroscopic quantum states as (3.30) in connection with the measurement problems is, however, slightly out of place.

Indeed, the scope of any experimental device is opposite: to acquire the information about the microscopic system ψ under consideration, and to record it in the form of a unique, classical state.

When $|\Psi\rangle$ represents a system with an infinite number of degrees of freedom, there is an exception to the possibility of superposition of (for instance) two, degenerate macroscopic states.

The system, at sufficiently low temperatures, will choose one of them as its ground states.

It could be tempting to relate such a phenomena of spontaneous symmetry breaking to the wave-function collapse.

For instance, one might try to construct a dynamical model of spin 1/2 particle, interacting with a doubly degenerate (2D) Ising model ground states, so as to mimic the state-vector reduction, (3.20).

More generally, the Ising model system may be replaced by some classical "pointer" states; the idea is to construct a model of nonunitary evolution of the microsystem ψ coupled to the pointer, to reproduce effectively (3.20), with correct relative frequencies for different outcomes.

Whether or not such a model of the wave-function collapse can be successfully constructed eventually, the conclusion from our discussion is the following:

No dynamical models of wave-function collapse are needed.

In our view, "wave-function collapse" is one of the worst misnomers in the quantum mechanics discussions.

The words evoke in our mind a mysterious, nonlinear evolution which shrinks almost instantaneously whatever distribution present in ψ before the measurement.

No such processes exist.

Every experimentalist knows what happens in each measurement: they are chain-ionizations and amplification, hadronic cascades in a calorimeter, recording of the particle tracks, etc.

They represent all nonadiabatic, irreversible processes.

However complicated in detail, they are all processes we understand well in principle, in terms of the standard theory of the fundamental interactions.

The state reduction (3.20) is just the way we perceive the fact that the result of each measurement is necessarily one of the terms in (3.7).

The wave functions of the states of the microsystem-device coupled system, corresponding to different measurements, have no common spacetime support, hence the wave function apparently - only apparently - "collapses".

Three puzzles

Before concluding, let us go through quickly three familiar puzzles and their resolution.

Quantum jumps versus Schrodinger's equation

A question often debated is how to reconcile the smooth time evolution of the wave function described by the Schrodinger equation with sudden "quantum jumps", occurring in radioactive nuclei (α and β decays), or in atoms in excited states, and so on.

Let us consider an α decay from a metastable nucleus, X,

$$\substack{A \\ N}(X) \Longrightarrow \substack{A-4 \\ N-2}(Y) + \alpha \, , \tag{4.1}$$

where N is the atomic number and A is the mass number.

The wave function for the system may be written as

$$|\psi\rangle = |X\rangle + |Y\rangle|\alpha\rangle \ . \tag{4.2}$$

We do not bother to write the coefficients in front of the two terms: their norms are not conserved due to the decay.

The first ($|X\rangle$) corresponds to a bound state; the second ($|Y\rangle|\alpha\rangle$) an unbounded system.

The interference between them is absent. (4.2) is not the wave function of a pure state.

It is a mixed state.

The state being metastable, its energy has a small imaginary part,

$$E = E_R - i\frac{\Gamma}{2} \ , \tag{4.3}$$

where Γ represents the total decay rate per unit time (or the level width), and $\tau = \hbar/\Gamma$ is the mean lifetime.

The wave function has a time dependence

$$\tilde{\psi}_X(t) = e^{-iEt/\hbar}\tilde{\psi}_X(0) = e^{-iE_Rt/\hbar}e^{-\Gamma t/2\hbar}\tilde{\psi}_X(0) \ . \tag{4.4}$$

Now how can one reconcile such a smooth time dependence with sudden quantum jumps, such as α decay, a spontaneous emission of photons from an excited atom?

This sort of question, probably mixed up with a philosophical confrontation between Heisenberg's view (based on the matrix mechanics, for the transition elements) and Schrodinger's one (with a smooth differential equations) of quantum mechanics, dominated the earlier debates on quantum mechanics.

Actually, from the very way the wave function $\tilde{\psi}$ for a metastable state is defined, where Γ represents the total decay rates of all possible decay processes of the parent particle, it is quite clear that $\tilde{\psi}$ represents an effective description of the metastable "quantum state", in which the coarse-grained time dependence (decay events) has been smoothed out.

It does not have the same status as the wave function of a genuine quantum state ψ.

In conclusion, the conundrum of apparent impossibility of reconciling the smooth time-dependent Schrodinger equation with quantum jumps, appears to have been caused by the confusion between the concept of the true wave function ψ and that of an effective "wave function" for a metastable "state", $\tilde{\psi}$.

Schrodinger's cat

In the discussion involving Schrodinger's cat, the initial "wave function" is a combination (4.2), between the undecayed nucleus $|X\rangle$ and the state after decay, $|Y\rangle|\alpha\rangle$.

To maximally simplify the discussion, let us eliminate altogether the intermediate, diabolic device which upon receipt of the α particle leads to the poisoning of the cat, and treat the cat directly as the measuring device, $|\Phi\rangle$.

When it detects the α particle, it dies; when it does not, it remains alive.

There are two terms in the measurement process, (3.6), (3.7), which here reads

$$|\psi\rangle \otimes |\Phi_0\rangle = (|X\rangle + |Y\rangle|\alpha\rangle) \otimes |\Phi_0\rangle \qquad (4.5)$$

$$\longrightarrow \quad |X\rangle \otimes |\Phi^{(alive)}\rangle + |Y\rangle|\alpha\rangle \otimes |\Phi^{(dead)}\rangle \qquad (4.6)$$

(we dropped also the rest of the world, |Env⟩).

Such a process appears to lead to the superposition of the dead and alive cat, which is certainly an unusual notion, difficult to conceive, to say the least.

Actually, there are some abuse and/or misuse of concepts in this argument, each of which individually invalidates it, eliminating the notorious conundrum.

The first is the fact, as seen earlier, that the "state" of the undecayed nucleus, X, is not a pure state.

The "wave function" $\tilde{\psi}_X$ describing it is not a proper wave function but an effective one, in which the coarse-grained time dependence has been averaged out.

Second, the linear superposition $|\psi\rangle \sim |X\rangle + |Y\rangle|\alpha\rangle$, as discussed around (4.2), does not represent a pure state, but a mixed state: $|Y\rangle|\alpha\rangle$ represents the decay product of $|X\rangle$, unbounded and incapable of interfering with the latter.

Finally, even putting aside these two issues, the process (4.5), (4.6) has clearly all the characteristics of a general measurement process discussed earlier.

As explained there, the wave functions describing the two terms of (4.6) lack overlapping spacetime supports.

There are no interferences between the terms involving the coupled system involving the microsystem and the classical measurement device (the cat here).

The formula (4.6) represents a mixture.

Equivalently, and perhaps more intuitively, one can use the notion of the wave-function collapse.

The exact timing of the spontaneous α decay cannot be predicted, as it reflects a quantum fluctuation.

But the moment an α particle is emitted, and the cat gets hit by it (the instant of the measurement), the wave function collapses,

$$|\psi\rangle \otimes |\Phi_0\rangle \Longrightarrow |Y\rangle|\alpha\rangle \otimes |\Phi^{(dead)}\rangle . \tag{4.7}$$

Recapitulating, Schrodinger's cat paradox, was caused by improper use of concepts such as the superposition principle, unitary and linearity of evolution, and by mixing up the pure and mixed states

What happens actually is simple: as long as the nucleus has not decayed (no α particle emission) the cat remains alive.

The exact timing of the α decay cannot be predicted.

But the moment the α particle is emitted, and the cat gets hit, it dies.

That is all.

There are no problems in describing this process appropriately in quantum mechanics, as seen above.

Setting aside these arguments, the idea of superposition of dead and alive cat states, represents in itself a self-contradictory, impossible notion.

A macroscopic superposition of states such as (3.30) or (4.6), requires the body temperature of the system close to T = 0, whereas a living cat needs room (body-) temperatures for its biological functions: it is necessarily a mixed state.

Quantum entanglement, quantum nonlocality and EPR paradox

Let us consider now the famous Einstein-Podolsky-Rosen setting (in Bohm's version) of a total spin 0 system decaying into two spin $\frac{1}{2}$ particles flying away from each other.

The (spin) wave function is given by

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle|\downarrow\rangle - |\downarrow\rangle|\uparrow\rangle) \qquad (4.8)$$

where $|\uparrow\rangle$ $(|\downarrow\rangle)$ represents the state of the 1st and 2nd spins, $s_{1,z} = \pm \frac{1}{2}$, $s_{2,z} = \pm \frac{1}{2}$.

As spin components of the single spins do not commute with the total spin $\mathbb{S}^2_{tot}$, e.g.,

$$[\mathbb{S}^2_{tot}, s_{1,i}] \neq 0 , \quad [\mathbb{S}^2_{tot}, s_{2,j}] \neq 0 , \qquad i, j = x, y, z , \qquad (4.9)$$

s_{1z}, s_{2y}, etc., are fluctuating between $\pm \frac{1}{2}$ in the state Ψ_0.

Note that Ψ_0 can be written in infinitely many different ways, reflecting possible fluctuation modes of the subsystems

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} (|+\rangle|-\rangle - |-\rangle|+\rangle) = \frac{1}{2} [|\uparrow\rangle(|+\rangle - |-\rangle) - (|+\rangle - |-\rangle)|\uparrow\rangle] = \dots . \qquad (4.10)$$

It is of fundamental importance to realize that such (correlated) fluctuations of the two subsystems - quantum entanglement - are present however distant the subsystems may be, and even if they are relatively spacelike separated.

Nothing in Ψ_0 tells us otherwise.

Nothing in quantum mechanical law says that such entanglements are present only when the subsystems are nearby, e.g., less than 10^{-8} cm, or less than 10^{+28} cm. Simply,

Quantum mechanics does not contain any fundamental parameter with the dimension of a length

In the EPR-Bohm experiment, this means that an experimental result at one arm, say, $s_{1z} = \frac{1}{2}$, might appear to imply "instantaneously" that the second experiment at the other arm would be in the state $s_{2z} = -\frac{1}{2}$, even before it is actually performed, and however distant they may be.

This could sound paradoxical (sometimes the term "quantum nonlocality" is used).

The second experimentalist, not having access to the first experiment, might have (should have?) expected to find the results $s_{2z} = \pm \frac{1}{2}$, a priori with equal probabilities.

This kind of argument has led to the introduction of the hidden-variable hypothesis, although such a deduction was not really justified.

We limit ourselves here to noting that this argument had logical flaws.

First, the contemporaneity of the two experiments is not really an issue: the experimentalists must just ensure that they are studying the two spin $\frac{1}{2}$ particles from the same decay event - the coincidence check.

The exact time ordering, which depends on the reference system chosen, cannot matter.

In any case, as the two subsystems are spacelike separated, it is untrue that the second experimentalist would know instantaneously the result of the first experiment, or actually, even whether or not the first measurement has been indeed performed.

The information transmission is itself a dynamical process.

Finally, the two "simultaneous" experiments would capture only those states of the two subsystems fluctuating according to the entangled wave function, (4.8), (4.10).

In conclusion, for the second experimentalist the system presents itself as a mixture with an unknown density matrix.

He (she) simply does not know what to expect.

Therefore there was no paradox, whatsoever.

It is true that the first experimental result $s_{1z} = \frac{1}{2}$ does imply, instantaneously (whatever it may mathematically mean), that the other spin is in the state $|\downarrow\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$, as is seen from the wave function (4.8), (4.10).

This quantum nonlocality, just a name for this particular aspect of quantum entanglement, is real.

It should, however, not be confused with the dynamical concept of locality (or causality).

Quantum mechanics is perfectly consistent with causality and locality, due to the fact that all the fundamental, elementary interactions are local interactions in spacetime.

No dynamical effects, including the information transmission, propagate faster than the velocity of light.

It is perhaps best to regard (4.8) simply as a particular kind of macroscopic quantum-mechanical state.

It is essential to realize that, unless the second measurement is done, any components of the second spin other than s_{2z} are still fluctuating.

This shows that the often mentioned "Bertlmann's socks" classical action-at-distance analogy, is not valid.

Quantum non-locality is subtler, and is different.

Thus if the two Stern-Gerlach type experiments are performed at the two ends simultaneously, by using the magnets directed to generic directions a and b, they would find, experiment by experiment, apparently random results, such as $(s_{1a}, s_{2b}) = (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, -\frac{1}{2}), (-\frac{1}{2}, -\frac{1}{2})$, etc.

But their fluctuation average is encoded in Ψ_0:

$$\langle \Psi_0 | s_a s_b | \Psi_0 \rangle = -\frac{1}{4} \cos \theta_{a,b} = -\frac{1}{4} \mathbf{a} \cdot \mathbf{b} . \tag{4.11}$$

A possible way to discriminate between any alternative theory with hidden variables from quantum mechanics, has been mathematically formulated, e.g., in the form of the Bell, or CSHS inequalities for analogous, polarization correlated photon pair experiments.

Beautiful experiments by Aspect et. al. (1981) have subsequently demonstrated that, whenever the hidden-variable alternatives and quantum mechanics give discrepant predictions (for certain sets of polarizer axes), the experimental data confirm quantum mechanics, disproving the former.

A reflection

As reviewed earlier, quantum correlations (entanglement) among the particles whose positions are spacelike separated, and hence are no longer capable to communicate with, or dynamically influence on, each other, as the two spin 1/2 particles in (4.8), (4.10), are quite a common, ordinary and ubiquitous phenomenon in QM.

What is not quite ordinary, is that this "quantum nonlocality" does not represent any violation of dynamical causality or locality (sometimes called Einstein causality).

Quantum mechanics is rigorously consistent with this principle, even though the proof of this fact requires working in the framework of relativistic quantum field theory, quantum theory of particles.

Now how can these two, apparently contradicting properties of quantum mechanics coexist peacefully within the same theory?

The answer lies in another fundamental aspect of QM: the wave function is not itself a measurable physical quantity (an observable).

Only the relative frequencies for different outcomes (probabilities) are predicted by it.

In hidden-variable models, the statistical aspect of quantum mechanics is replaced by a classical statistical (unknown) distribution of the hidden parameters $\{\lambda\}$, complementing the standard description with the wave function.

But once the value(s) of $\{\lambda\}$ at $t = 0$ is (are) chosen, all possible imaginable future experimental outcomes are determined uniquely (i.e., classical evolutions).

A hidden-variable theory, with a logical structure so sharply different from that of quantum mechanics, cannot reproduce all of the predictions of the latter.

It could, actually, if one is willing to introduce nonlocal, noncausal evolutions.

This is seen clearly, e.g., in the unphysical, noncausal behavior of the trajectory $X(t)$ in Bohm's pilot-wave theory which is a particular type of hidden-variable model, designed by construction to reproduce all QM predictions.

If we insist, instead, that a hidden-variable model must respect local causality, then its predictions necessarily satisfy certain inequalities.

Bell and Clauser et. al. have shown that some predictions of QM should lie outside such a bound, a fact verified, in a dramatic experimental confirmation of QM.

Conclusion

We have now answered the "quantum measurement problems" enlisted in the beginning of our discussion.

The key observation of this discussion that the basic entities (the degrees of freedom) of our world are various types of particles, leads to the idea that each measurement, at its core, is a spacetime, pointlike event.

The crux of the absence of coherent superposition of different terms in the measuring process (3.7) is, indeed, the consequent lack of the overlapping spacetime supports in the associated wave functions, instantly after measurement which is the process of entanglement between the microsystem and the experimental device and the environment.

It is in this sense the expression (3.7) represents a mixed state.

But it also implies that the aftermath of each measurement corresponds to a single term of (3.7).

This (also empirical) fact, is known as the "wave-function collapse", or the state-vector reduction.

The discussions have led to the diagonal density matrix, (3.18).

In particular, the information encoded in the wave function $|\psi\rangle = \sum_n c_n|n\rangle$ has been transferred via entanglement with the experimental device and environment into the relative frequencies for finding the experimental results $F = f_n$, $\mathcal{P}_n = |c_n|^2$.

Combined with the equality between $\mathcal{P}_n = |c_n|^2$ (the relative frequencies) and $P_n = |c_n|^2$ (the probabilities) our results represent the first steps towards filling in the logical gaps in Born's rule, by explaining the state-vector reduction, and pointing towards a more natural interpretation of quantum mechanics.

Our discussions do confirm the standard Born rule as a concise way of summarizing the predictions of quantum mechanics.

But the fact that the latter is formulated in terms of the "probability certain result is found in an experiment, which might (or might not) be eventually performed - perhaps by an experimentalist with a Ph D - has always caused confusion and conceptual difficulties.

Every teacher of a quantum mechanics course knows about this curious feeling of guilt he (or she) experiences, on the day of proclamation of the "fundamental postulate" (Born's rule).

   In ultimate analysis, the problem of Born's rule is the fact that it presumes a human intervention.

We instead take the quantum fluctuations described by the wave function as real, and propose that they represent the fundamental laws of quantum mechanics.

They are there, independently of any experiments, and in fact, whether or not we human beings are around.

What we propose here is a slightly unconventional way of understanding quantum mechanical laws.

Indeed, the solid textbook materials and the standard methods of analysis in quantum mechanics, all remain unchallenged.

After all, we know that quantum mechanics is a correct theory.

The new perspective, however, suggests us to dispense with, in the discussion of quantum mechanics, certain familiar concepts such as probability, statistical ensembles, Copenhagen interpretation, etc., which historically have guided us, but also caused misunderstandings and introduced clouds in the discussion on quantum mechanics, for many years.

Accepting the quantum fluctuations described by the wave function as real, we simply give up pretending, or demanding, that in a fundamental, complete theory describing Nature the result of each single experiment should necessarily be predictable.

As cosmologists tell us today, all the known structures of our universe (including galaxies, stars, planets and ourselves) are believed to have grown out of uncontrollable quantum density fluctuations at some stage of the inflationary universe. In other words:

God may not play dice, but Nature apparently does.

We believe that our new, clearer understanding the quantum mechanical laws of Nature, has far-reaching consequences and applications in all present and future science domains in physics and beyond, from biology and chemistry, quantum information and quantum computing, and possibly, to the brain science.