



Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

Введение в профессиональную деятельность

Лекция 7

Свойства реальности, выраженные с помощью данных. Что такое «наука о данных?»

27 марта 2018 г.

Что обсуждали на прошлой лекции

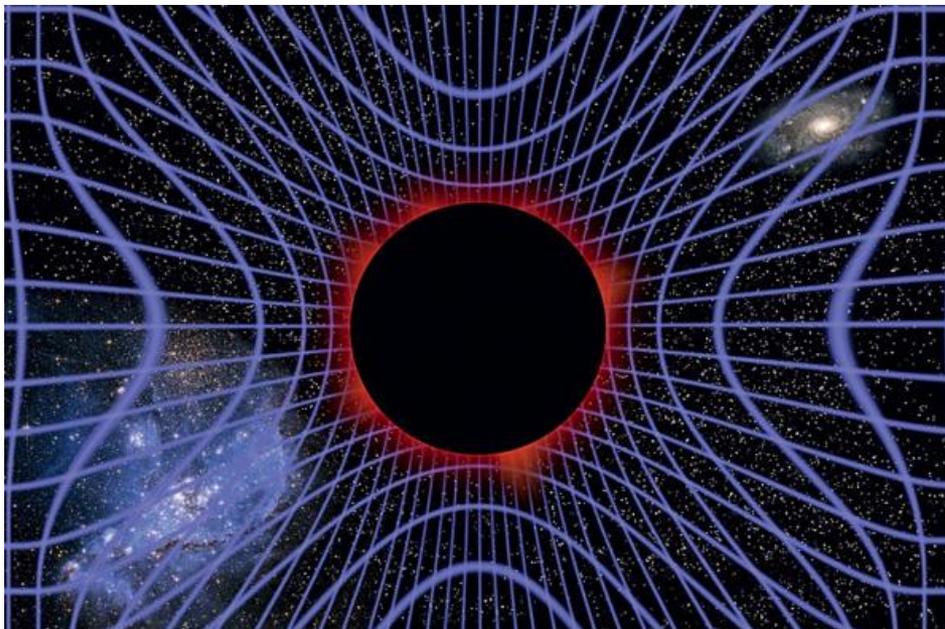
Из концепции **Digital physics** (Кónрад Цúзе, книга **Calculating Space**) следует вывод, что физическая реальность на фундаментальном уровне **вычислима**, т.е. обладает свойствами, которые непосредственно выражаются через количественные отношения, которые, в свою очередь, изоморфны программам компьютерных вычислений.

Но кроме «физической» существует также и «метальная реальность». Вычислима ли она ?

Ключевой вопрос компьютерных наук: Можно ли **любой** процесс, носителем которого является окружающая человека реальность, «**ВЫЧИСЛИТЬ**»?

Почему абстрактные математические теории помогают понять, как устроен наш мир ?

Современная физика «прежде всего» старается угадать математический аппарат, оперирующий данными, о свойствах которых заранее вообще ничего не ясно, что они означают



В законе Ньютона фигурируют силы тяготения, которые можно измерять непосредственно, и определяются они однозначно (в выбранной системе единиц). А вот значения гравитационного потенциала можно изменить на любую постоянную величину — градиент останется тем же.

Проблема «данных» И Data Processing

- Одному и тому же физическому полю могут соответствовать разные потенциалы. Например, к векторному потенциалу можно добавить любой постоянный вектор, а к скалярному — любое число
- Наука о данных (англ. data science или «дательгия» — datalogy) — раздел компьютерных наук, изучающий проблемы анализа, обработки и представления данных в цифровой форме в зависимости от контекста их получения.
- Метафора Dig Data — это символ надвигающейся технической революции а вместо термина Information Technology следовало бы употреблять более точный термин Data Processing

История вопроса

- Первый опыт в Data Processing датируется IV тысячелетием до нашей эры, когда появилось пиктографическое письмо.
- С тех пор самое мощное представление данных – это текстовое представление.
- Различного рода математические численные методы от теоремы Пифагора и табличных приемов упрощения расчетов до современных компьютеров – это цифровые данные. Они должны быть «точными», чтобы результаты расчетов соответствовали физическим процессам. Точность требует большого числа разрядов и времени расчетов.
- Рост объемов цифровых данных вместе с необходимостью их анализа **в режиме реального времени** требуют создания и внедрения новых инструментов, позволяющих эффективно решать задачу Big Data Analytics – извлечения из данных «знаний»

Что такое «наука о данных» ?

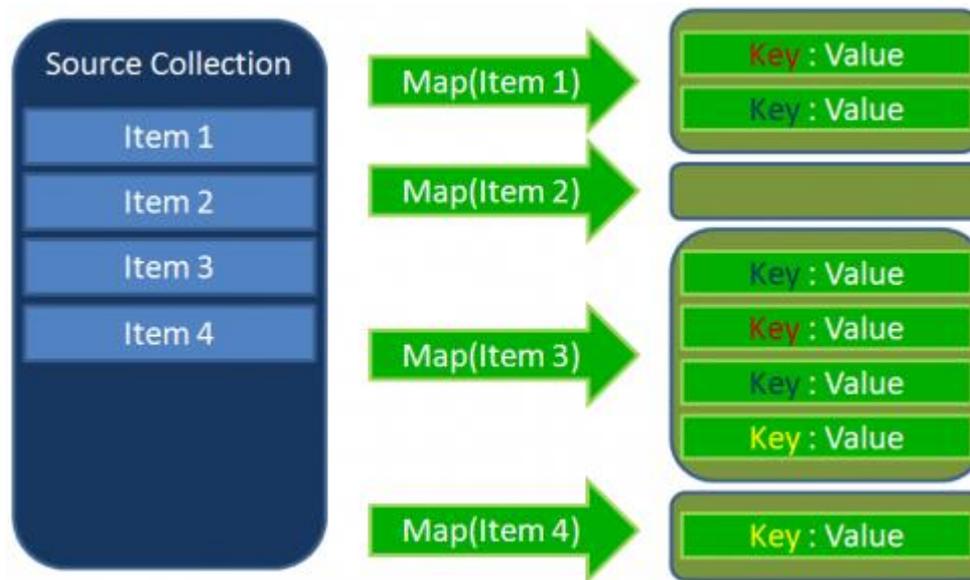
- Для анализа больших объемов данных применяются различные методы «математики больших данных»:
 - нейронные сети — модели, построенные по принципу организации и функционирования биологических нейронных сетей;
 - методы предсказательной аналитики,
 - статистики и Natural Language Processing (направления искусственного интеллекта и математической лингвистики, изучающего проблемы компьютерного анализа и синтеза естественных языков).
 - методы, привлекающие экспертов, или краудсорсинг

Новые подходы

- NoSQL базы данных и СУБД не подразумевают внутренних связей, использование «ключ-значение» - хеш функция
- СУБД Cache (произносится: «кашэ») — иерархическая СУБД, позиционирующаяся мультимодельная, т.е. одновременный доступ к данным как «объекту» (структуру данных) или «реляционной сущности» (структура данных как таблица).

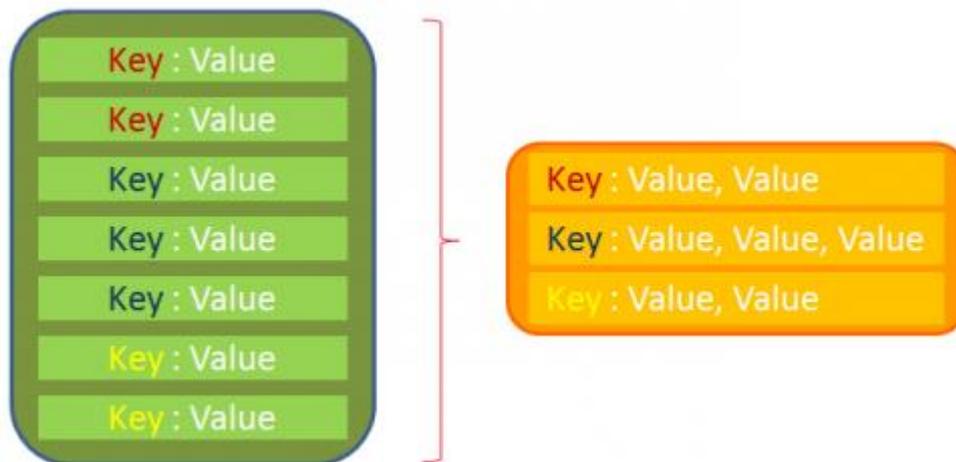
Работа с «Большими данными» - MapReduce

- MapReduce — модель распределённых вычислений, используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт, наборами данных в компьютерных кластерах.



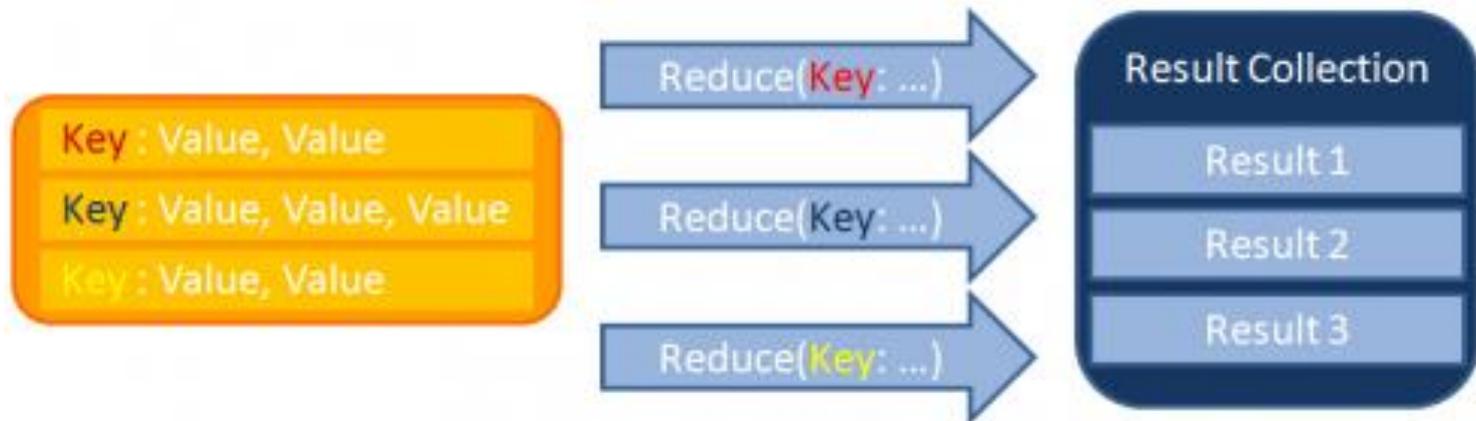
Классификация объектов по «ключам»

Новые экземпляры объектов, где все значения (value) сгруппированы по ключу.



Функция Reduce

Возвращает новый экземпляр объекта, который включен в результирующую коллекцию



Данные как атрибут

- Физическая реальность это совокупность объектов и процессов (изменения объектов), с которыми мы судим на основе данных.
- К объектам физической реальности применим фундаментальный принцип "**тождества неразличимых**", сформулированный Готфридом Лейбницем.



- Суть этого принципа в том, что любые два физически **неразличимых** объекта (субстанции) неизбежно совпадут, став тождественным объектом (субстанцией).
- Возможность различать одни объекты природы от других неизбежно требует, что объекты являются носителями некоторой меры разнообразия, т.е. *difference that make a difference** - это и есть **информация**

Если компьютеры – физические устройства, а вычисления информационный процесс, то надо знать:

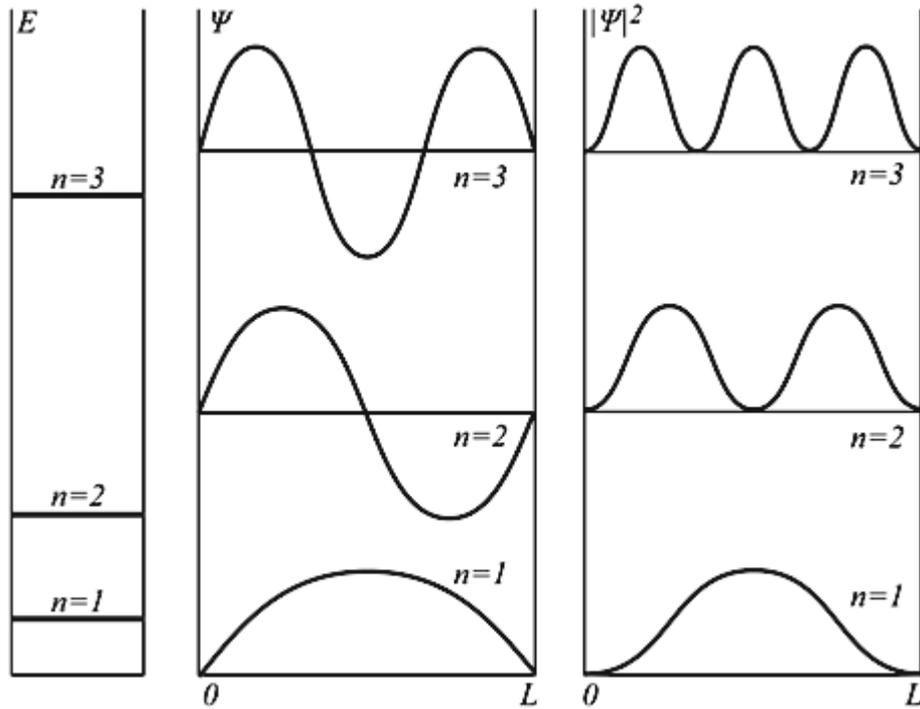
- Какие **законы** физики лежат в основе процессов обработки данных ?
- Что есть **результат** процесса вычислений – новый физический процесс, новый объект или новые данные ?
- Можем ли мы организова «правильную» последовательность вычислений смоделировать и даже «формировать» саму реальность как совокупность данных ?!!!
- Что является **мерой** точности результата вычислений - цифровых данных ?

Предварительные выводы

- Мир это совокупность информационных и физических процессов.
- «Носителем» процессов является данные и средства их хранения память – механизмом переноса состояний во времени.
- Рассматривая данные как атрибут реальности, можно предложить «формулу» физической реальности :

материя= (вещество + энергия) + данные .

Физика «волновых функций» состояния



Энергия определяет «частоту» волновой функции квантовой частицы Ψ_i

Квадрат модуля волновой функции $|\Psi|^2$ определяет вероятность нахождения частицы в i -ом состоянии

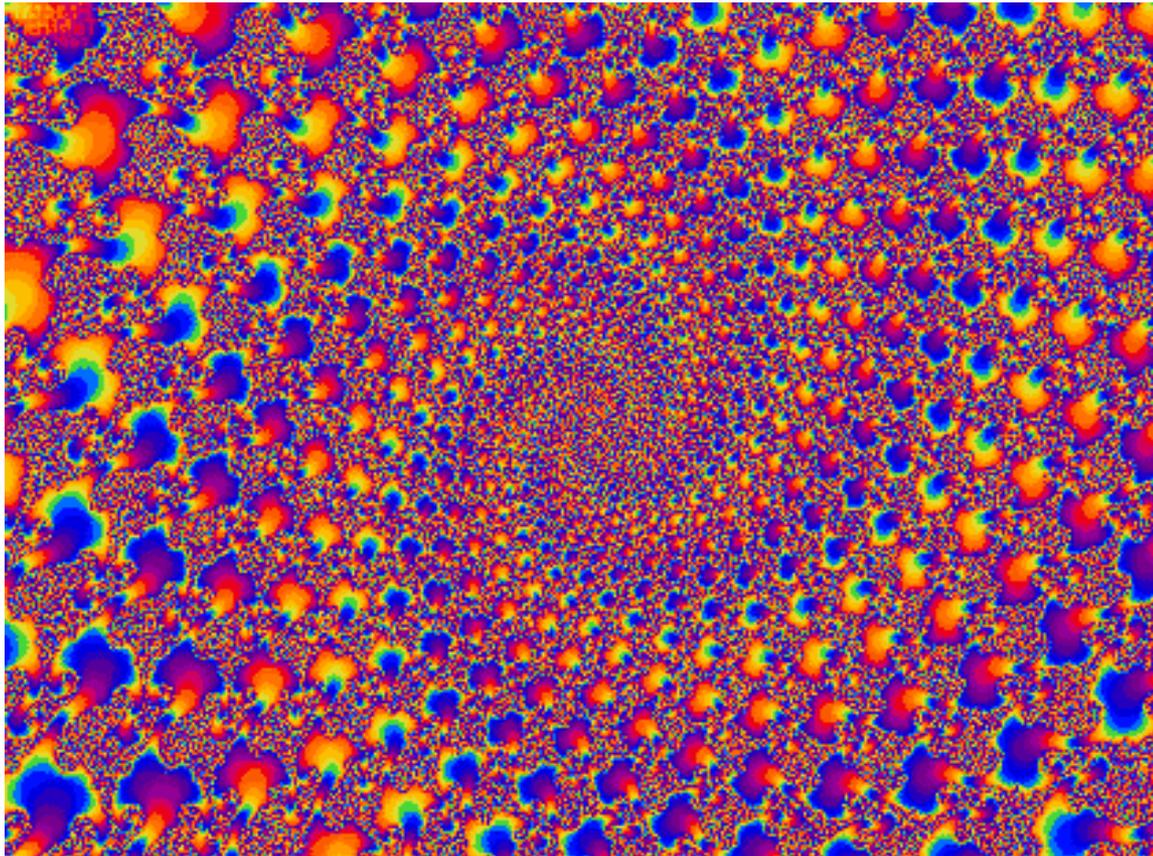


Физическое «запутывание» возможных состояний атома и внешнего силового поля



- Различные физические понятия, такие как поля или частицы, с позиций информационного формализма рассматриваться лишь как разные способы количественной характеристики целостной, непротиворечивой и объективно существующей физической реальности
- В метафорической форме идею информационного описания реальности сформулировал лауреат Нобелевской премии выдающийся физик 20 века Дж. А. Уильер - «it from bit»,

Фрактал – пример самоподобного вездесвязанного множества



Почему нужна «новая физико-информационная модель реальности »:

Существуют явления, которые нельзя объяснить в рамках классической физики:

- поведение косяка мелких рыбешек или мелких птиц перед их отлетом на юг;
- поведение толпы людей на концерте популярного ансамбля;
- явления предшествующие землетрясению: поведение животных, заряженные частицы в атмосфере, свечение воды, образование облаков и пр.;
- кооперативные явления в природе и обществе, самоорганизация, фликкер-шум и пр.;
-

Макроскопическая информационная «ЗАПУТАННОСТЬ» данных



Случай политическая
«полигамия»



случай
классическая
«моногамия»:



Topological order and quantum entanglement

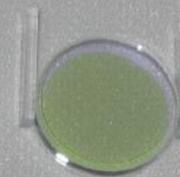
Michael Levin, Xiao-Gang Wen

MIT

<http://online.itp.ucsb.edu/online/qubit06/levin/>

LIGO

MIT Radiation Pressure Experiment Optical Coatings: Noise and Other Issues



Tim Bodiya

MIT: S. Ackley, T. Corbitt, D. Sigg, N. Smith, C. Wipf, N. Mavalvala
Caltech: Y. Chen, R. Bork, J. Heefner, S. Whitcomb, S. Danilishin
AEI: H. Ebhardt-Mueller, H. Rehbein, K. Somiya
LIGO Hanford Observatory: D. Sigg

Workshop on Optical Coatings in Precision Measurements
March 20-21

<http://www.ge.infn.it/~gemme/virgo/CoatingWorkshop/bodiya.pdf>



Quantum Information Technology: Entanglement, Teleportation and Quantum Memory

October 23, 2002

Applications of Entanglement

Seth Lloyd

MIT Department of Mechanical Engineering

Уточнения:

«Чистым» информационным состоянием называется такое состояние объекта, которое может быть описано вектором состояния.

«Несепарабельным» или **перепутанным состоянием** называют такое состояние составной системы, волновую функцию которого нельзя представить в виде тензорного произведения волновых функций составляющих ее частей A, B, \dots ,

$$|\Psi\rangle \neq |\Psi\rangle_A \otimes |\Psi\rangle_B \otimes \dots$$

Вывод: Если волновая функция объекта может быть представлена в виде такого произведения тензорного произведения, это означает, что система не содержит никаких корреляций – ни классических, ни квантовых

Пример перепутанного состояния - пара спинов $|\Psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow, \downarrow\rangle - |\downarrow, \uparrow\rangle)$.

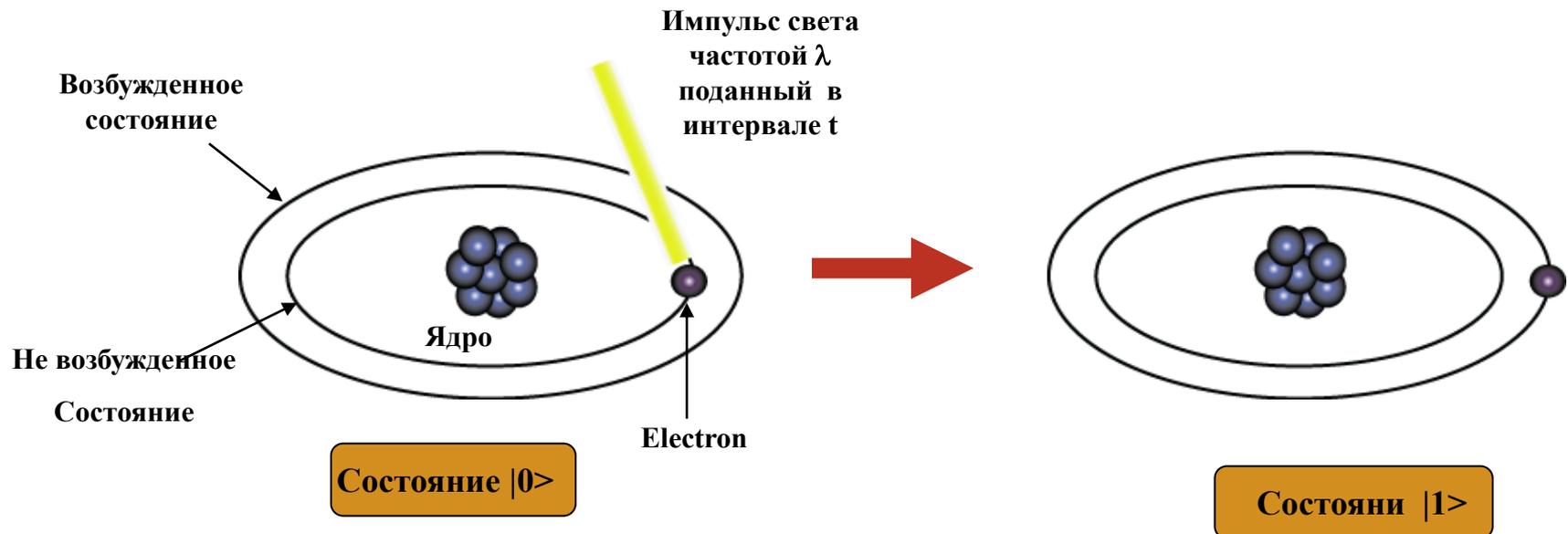
Пример носителя «особого состояния» - квантовый бит или Qubits

Чистое состояние атома, имеющего два различных энергетических уровней, может быть вектором. «Возбужденное» состояние обозначим вектором

$$|1\rangle = (0, 1)^T,$$

а не возбужденное вектором

$$|0\rangle = (1, 0)^T$$



Суперпозиция «несовместных состояний»

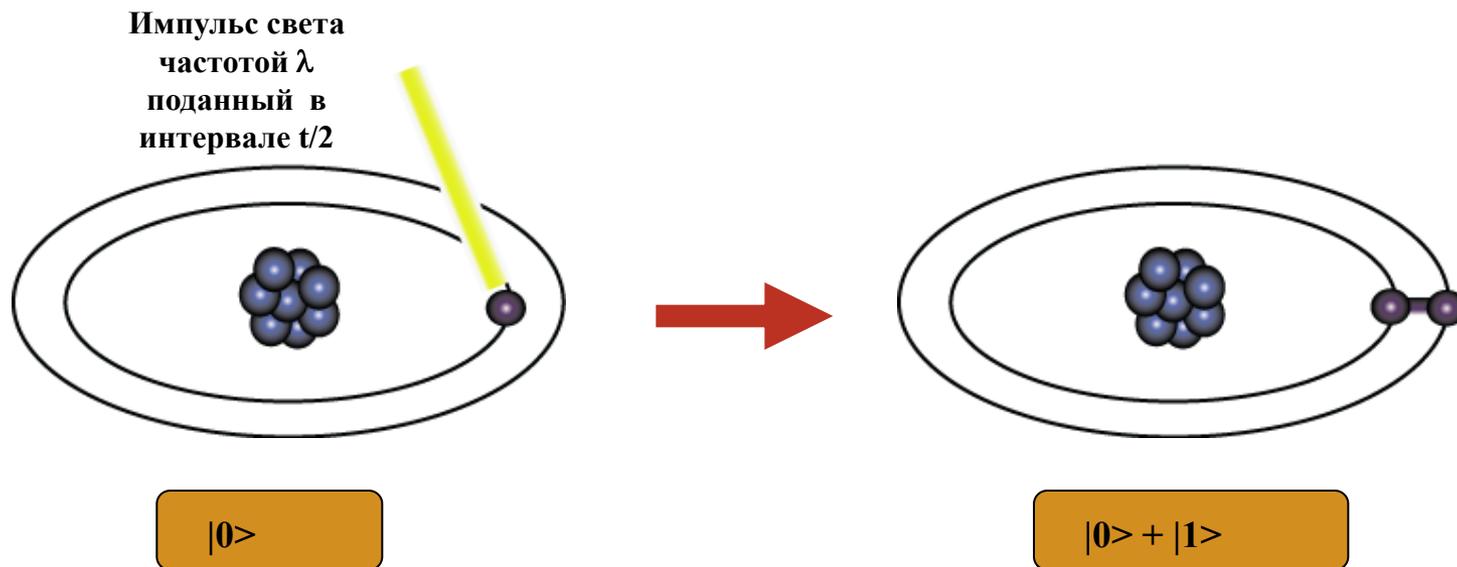
Атом как объект реальности описывается вектором или «волновой функцией»

$$|\psi\rangle = \alpha_1 |0\rangle + \alpha_2 |1\rangle$$

где α_1 and α_2 комплексные числа , которые еще и нормированы

$$|\alpha_1|^2 + |\alpha_2|^2 = 1$$

Как возникает суперпозиция



В случае, когда кубитов 3, то суперпозицию всех 8 возможных состояний можно записать как:

$$|\psi\rangle = \frac{1}{\sqrt{8}} |000\rangle + \frac{1}{\sqrt{8}} |001\rangle + \dots + \frac{1}{\sqrt{8}} |111\rangle$$

Эволюция вектора состояний как информационный процесс

Оператор U порождает «движение» в соответствии с формулой :

$$|\phi(t)\rangle = U|\phi(0)\rangle,$$

где U отвечает уравнению Шредингера

$$i\hbar \frac{\partial}{\partial t} |\phi\rangle = H |\phi\rangle \quad U = \exp\left(\frac{i}{\hbar} \int H dt\right)$$

- Физика запрещает передачу данных «в пространстве-времени» со сверхсветовой скоростью.
- Вероятностная природа измерений (результат того, куда перейдет суперпозиция – случаен) открывает новые возможности информационного описания реальности.
- Стабильные информационные связи могут существовать как на квантовом уровне (в «пространстве-времени» так и на макроскопическом уровне (информационно-вычислительное пространство)