



Введение в профессиональную деятельность

Лекция 3

**Тема: Логика формальная vs.
разумного – что можно
вычислить помощью
компьютера ?**

Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

20, 27 февраля 2018 г.

Вместо введения

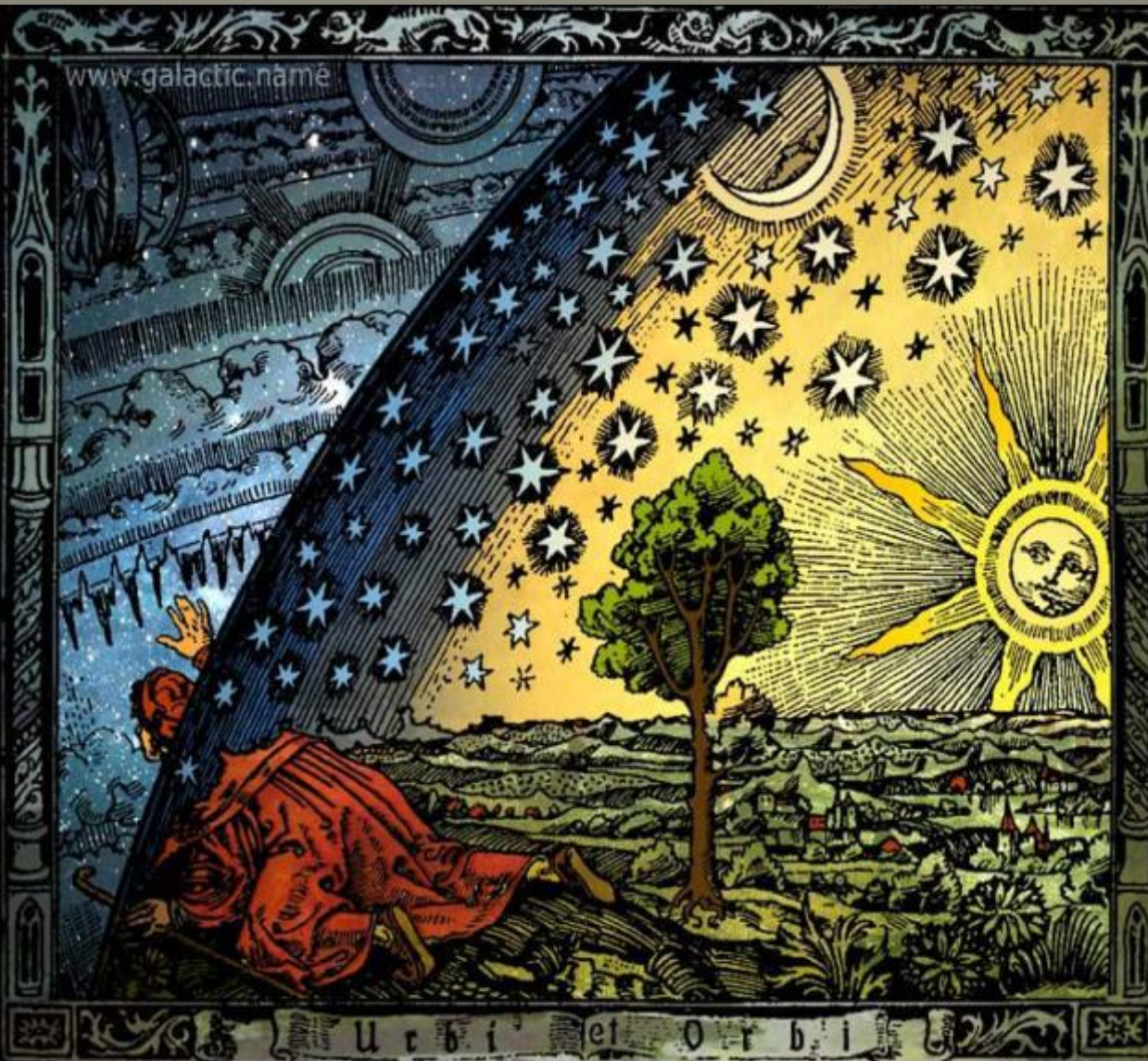
*Thinking is not the ability to manipulate language; it's the ability to manipulate **concepts**. Computer science should be about concepts, not languages.*

Лесли Лэмпорт

Выводы из прошлой лекции

- Мир – **сложная открытая система**. Феномен **информации** – феномен возможного, но еще не состоявшегося, связан с т.н. натуральными **вычислениями** и компьютерным моделированием
- Феномен **сложности** проявляется в форме суперпозиции **«потенциально несовместных состояний»** - для полного описания явлений необходимо применять два **взаимоисключающих** («дополнительных») **набора понятий**, совокупность которых даёт исчерпывающую информацию об этих явлениях как о целостных.
- В макром мире носителем «дополнительности» выступает сознание человека, которое за счет когнитивных способностей мозга может осуществлять **преобразование «возможное – состоявшееся»**.

«Потенциально возможное»



ищем ответ на вопрос:

Существуют ли математические операции, которые позволят построить «количественные» модели «потенциально возможного»?

Что такое «математика больших данных»?

гравюра Фламмариона

Границы реальности



Привычное представление о наличии одного и только одного исхода у любого эксперимента — в корне неверно, т.е. закон исключенного третьего не является универсальной основой научного **понимания** мира

$$A \vee \bar{A} = 1$$

«Цифровая» концепция будущего



Роботы и люди скоро станут «жить» в одном информационном пространстве

Что об этом уже известно и как мы оцениваем будущее

- 9.02.2016 г. *“Искусственный интеллект вытеснит с полей комбайнеров и трактористов. Курс на замещение водителей роботами”*

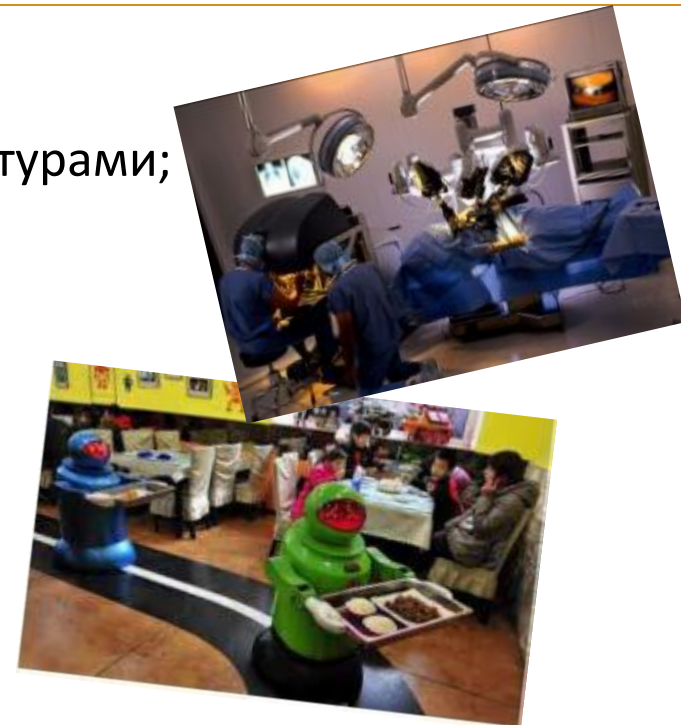
- 25.01.2018. *“Согласно исследованиям компании McKinsey, к 2030 году потеряют работу 800 млн чел.”*



- RoboCup: *“На чемпионате мира 2050 г. команды автономных роботов обыграют футбольные команды из людей”*

Мы вступаем в «цифровую эпоху» - computo ergo sum

- “Цифровизация” экономики;
- компьютерное управление критическими структурами;
- киберфизические системы;
- роботизированные производства;
- электронные правительства;
- роботы-хирурги;
- боевые роботы;
-



очень скоро люди будут **ВЫНУЖДЕННЫ**
сосуществовать с чужим интеллектом – с роботами,
которых они сотворят сами!

“И создал Господь человека по своему образу и подобию...”



Сотворение
Адама
Микеланджело
Сикстинская
капелла

Студентами, освоившим инструменты компьютерных наук, – жизнь даст уникальную возможность стать **ТВОРЦАМИ** новой **ИСКУССТВЕННОЙ ЖИЗНИ**, которая будет сосуществовать и **ЭВОЛЮЦИОНИРОВАТЬ** вместе с нами **на Земле и в космосе**

Что нас ждет

- *“Недалеко то время, когда робот осознает себя как некую индивидуальность.*
- *У него появятся свои планы, свои мотивы, свои цели, и, я вас уверяю, мы не будем входить в этот смысл.”*

Профессор
Наталья Черниговская

“Мы привыкли считать себя венцом творения. Оказывается, что это не так, на вершине эволюции будет человекомашина, кибернетический организм, киберфизические системы.”

Евгений Кузнецов

Но есть проблемы .. Ошибки при создании ПО неизбежны!?

Проблема в том, что

системы, которые мы строим, ведут себя в точности так, как мы их запрограммировали,

НО

совсем не так, как мы предполагали!

Человеческий мозг сам по себе не в состоянии справиться со сложностью при разработке программ объемом в сотни миллионов строк кода

“COMPUTERS ARE GOOD AT FOLLOWING INSTRUCTIONS, BUT NOT AT READING YOUR MIND.”

- DONALD KNUTH -



Норберт Винер

“Компьютеры - безотказные исполнители”

*“Если магия вообще способна даровать что-либо, то она дарует именно **то, что вы попросили, а не то, что вы подразумевали**, но не сумели точно сформулировать.*



Н.Винер: в компьютерах *“исполнение заданного осуществляется машинами в высшей степени буквально.”*

Всеобщий цифровой учет и контроль.

*«Компьютеры похожи на ветхозаветных богов:
сплошные правила и ни капли жалости».*

(Дж. Кэмпбелл)

***... и Если компьютер решит, что вы умерли, то вы
умерли ...***

Кто отвечает за ошибки в программах ?!

Обычно ПО имеет 10-15 ошибок на 1000 строк кода,
ПО высокого качества – 3 ошибки на 1000 строк кода.

Современное ПО, управляющее реальной технической системой,
состоит из сотен миллионов строк кода.

**Программы, управляющие техникой вокруг нас,
наполнены ошибками.**

Точное следование роботами инструкций приводит к авариям? (АиФ, 04.10.2016)

- Tesla начала продажи Model S с встроенным роботом, однако автономный Tesla улетел под прицеп разворачивающегося грузовика. Водитель погиб на месте. Причиной ДТП стало ослепление датчиков прямыми лучами солнца. *Компьютер не смог понять, что за объект вырос под носом.*
- Эксперименты Google и Jaguar Land Rover показали, что роботизированная машина, едущая исключительно по правилам, способна создать еще большую опасность, чем лихачи.
- К примеру, автопилот пока не умеет пропускать кареты скорой помощи или полицию, так как приходится вырывать в межполосное пространство.
- Автопилот не видит разницы между картонной коробкой и бетонным блоком. **Следуя машинной логике, он никогда не покинет кругового перекрестка, ведь для этого требуется формальное нарушение правил.** ... На круге действуют джентельменские соглашения, не понятные электронному разуму. Компьютер не распознаёт вежливости попутчиков

Интеллектуальные агенты и многоагентные системы

Новая область интеллектуальных систем называется "*многоагентными системами*", *МАС*

Метафора МАС – новый взгляд на мир, состоящий из агентов, в котором мы сами – только часть, биологические агенты.

Искусственные агенты – РОБОТЫ - скоро будут функционировать в обществе наряду с нами.

Разработка многоагентных систем требует качественно новых подходов, теорий и технологий программирования.

Каким он будет, наш мир с роботами?

Новые проблемы:

1. Как построить адекватную модель поведения?
2. Как организовать согласованное целесообразное взаимодействие коллектива агентов?
3. Как проверить, что построенная программная система эффективно решает делегированные ей задачи?

Трудно быть Богом. Историческая ответственность программистов



ГЛОБАЛЬНЫЙ ВЫЗОВ ЧЕЛОВЕЧЕСКОЙ ЦИВИЛИЗАЦИИ

- Мы сами создаем искусственный цифровой мир. Мы сами реализуем в нем законы поведения искусственных агентов.

Итого

Искусственные интеллектуальные агенты должны функционировать в соответствии с нормами, принятыми в человеческом обществе

Основной вопрос

Как формализовать нужные законы *поведения и этики* и внедрить их в “мозги” (ПО) агентов?

Как можно построить “**компьютерные мозги**”, которые будут взаимодействовать с ДРУГИМИ мозгами, машинными либо биологическими? Какие права и обязанности могут они иметь?



Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

Лекция 4

**нормативная логика
– логика законов и правил
поведения интеллектуальных
«агентов»**

27 февраля 2018 г.

Программисты боятся формальных методов

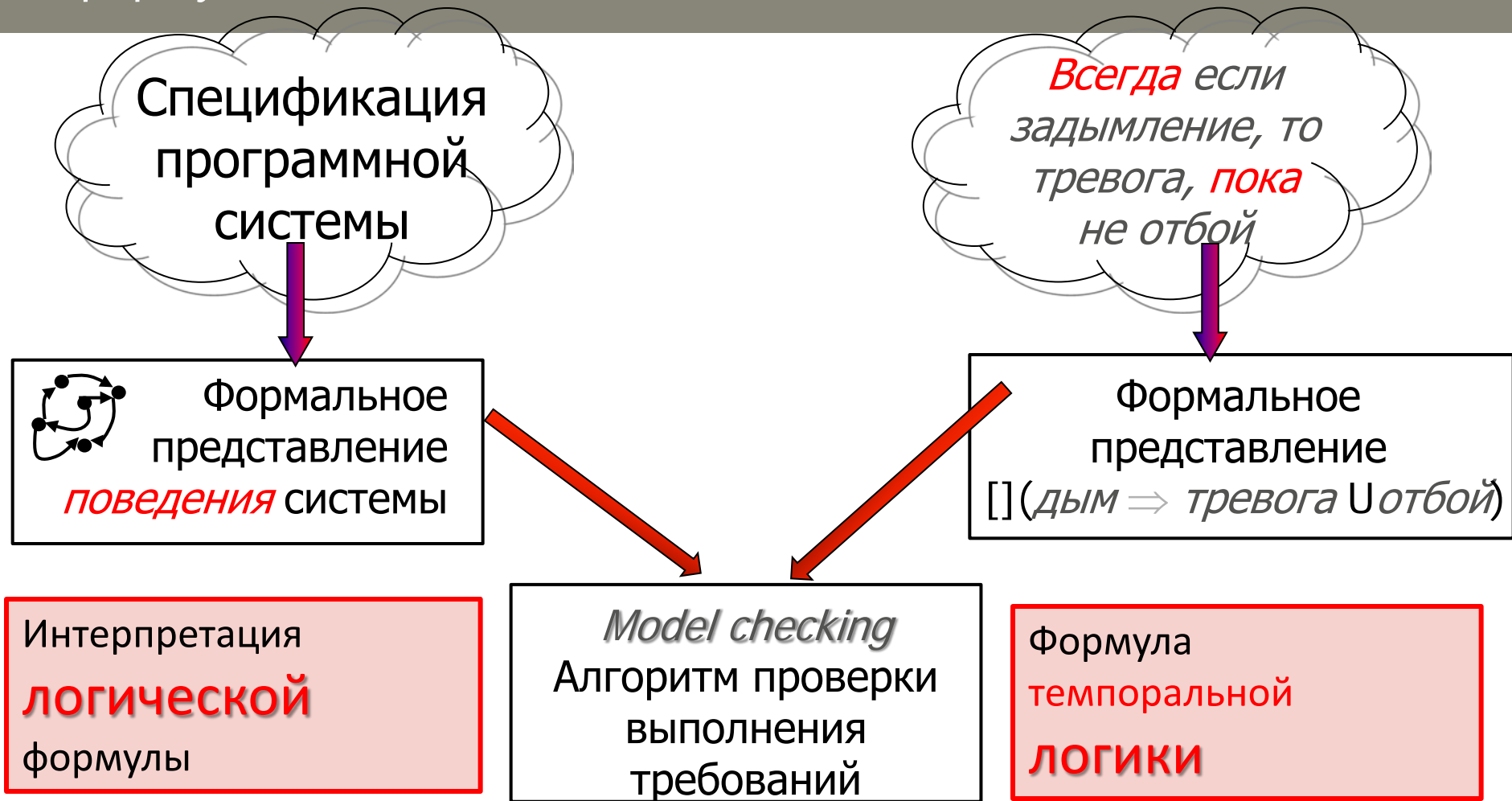
- Программисты строят свои программы, исходя из “здорового смысла”.

Разработать интеллектуальную многоагентную систему
с предсказуемым и контролируемым поведением,
исходя из “общих соображений”,

НЕВОЗМОЖНО !

Например, тонкие вопросы изменения знаний агентов, задания норм поведения агентов в социуме и множество подобных проблем требуют тонких формальных методов их спецификации и анализа.

Требования к БУДУЩЕМУ поведению программных систем – формулы ТЕМПОРАЛЬНОЙ логики



ВСЕ алгоритмы управления должны проверяться относительно их БУДУЩЕГО поведения во времени

Что такое модальность?

- ЭТО **характеризация** истины

- Иван  пылесосит.

обязательно:

Чжен знает, что;

в будущем будет;

случалось, что;

должна;

не имеет право ;

может

... ..

- Модальные операторы формально обозначаются:

[] "Иван _пылесосит"

или \diamond "Иван _пылесосит"

Модальная логика – любая логика, позволяющая формализовать и анализировать рассуждения с различными характеристиками ИСТИНЫ: возможность/необходимость, знание/вера и т.п.

Модальная логика – логика с модальными операторами

- **Модальность** (от лат. *modus* – вид, способ, наклонение) – *оценочная категория*, определяющая *отношение* высказывания к действительности. Модальная логика – любая формальная логическая система, в которой присутствуют модальные операторы.
 - $\diamond p$ – “*p* возможно (может быть) истинно”
 - $\Box p$ – “*p* обязательно (должно быть) истинным”

 - $\diamond p$ – “когда-нибудь в будущем *p* будет истинным”
 - $\Box p$ – “*p* будет истинным всегда в будущем”

 - $\Box_A p$ – “агент *A* **знает**, что *p* истинно”
 - $\diamond_A p$ – “агент *A* **предполагает**, что *p* истинно”
 - \Box_i “идет дождь” – “агент *i* **знает**, что идет дождь”.
- } *Логика*
возможности/
необходимости
- } *Временная*
логика
- } *Логика знаний*

Модальная логика: логика “сильных” и “слабых” истин

- МЛ – формализм для изучения *логически связанных пар модальностей* типа “необходимо” и “возможно”, которые могут стоять перед утверждениями (высказываниями) с возможными значениями “истина” или “ложь”.
Формально эти модальности обозначаются операторами \square и \Diamond .

Логика	Модальность необходимого $\square p$	Модальность возможного $\Diamond p$
Алетическая alethic	p необходимо (p is necessary)	p возможно (p is possible)
Докастическая doxastic	p обязательно (p is believed)	P предположительно (p is supposed)
Темпоральная temporal	всегда в будущем p (p always)	иногда в будущем p (p sometimes)
Деонтическая deontic	p должно быть выполнено (p is obligatory)	p разрешается выполнить (p is allowed)
Эпистемическая epistemic	знаю, что p (p is known)	думаю (полагаю), что p (p is believed)
...

модальные логики мультиагентных систем

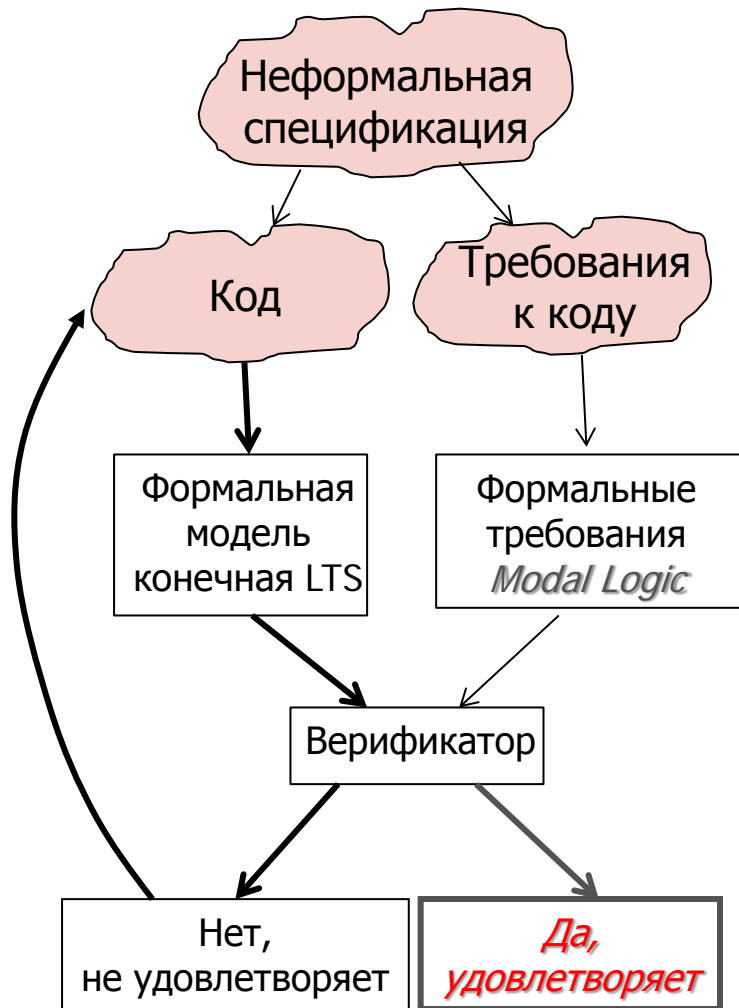
Формализация в области разработки многоагентных систем играет ту же роль, которую она играет в других областях инженерной деятельности:

*она позволяет изучать свойства разрабатываемой системы **на модели** вместо того, чтобы ликвидировать последствия некорректного функционирования системы, которая была изначально построена из “общих соображений”, на основе*

*“здравого смысла”. **Здравый смысл не помогает строить корректные системы!***

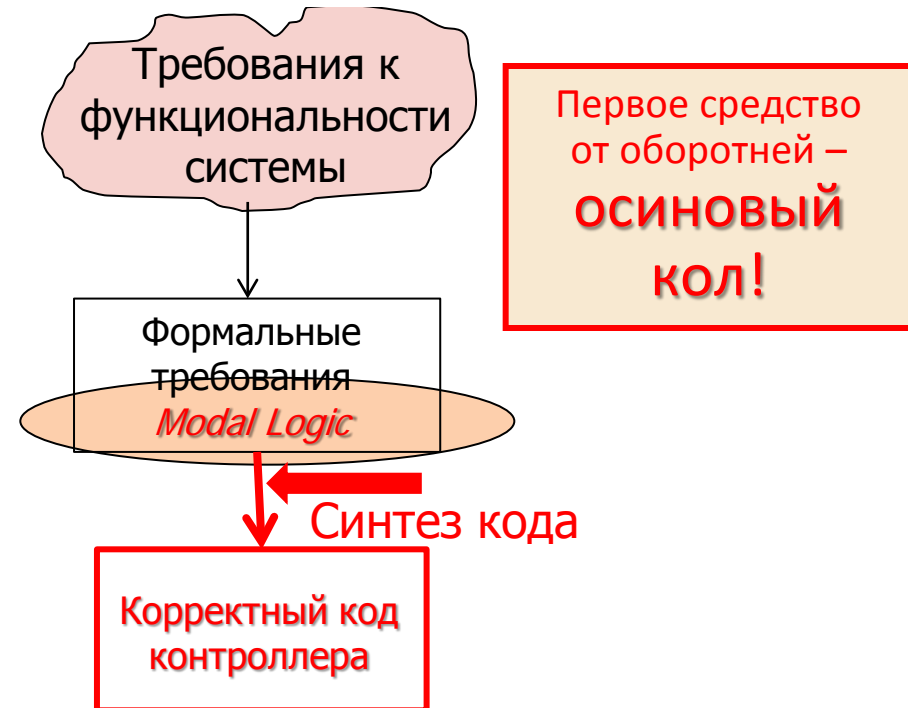
Нет «серебряной пули».

Классический подход



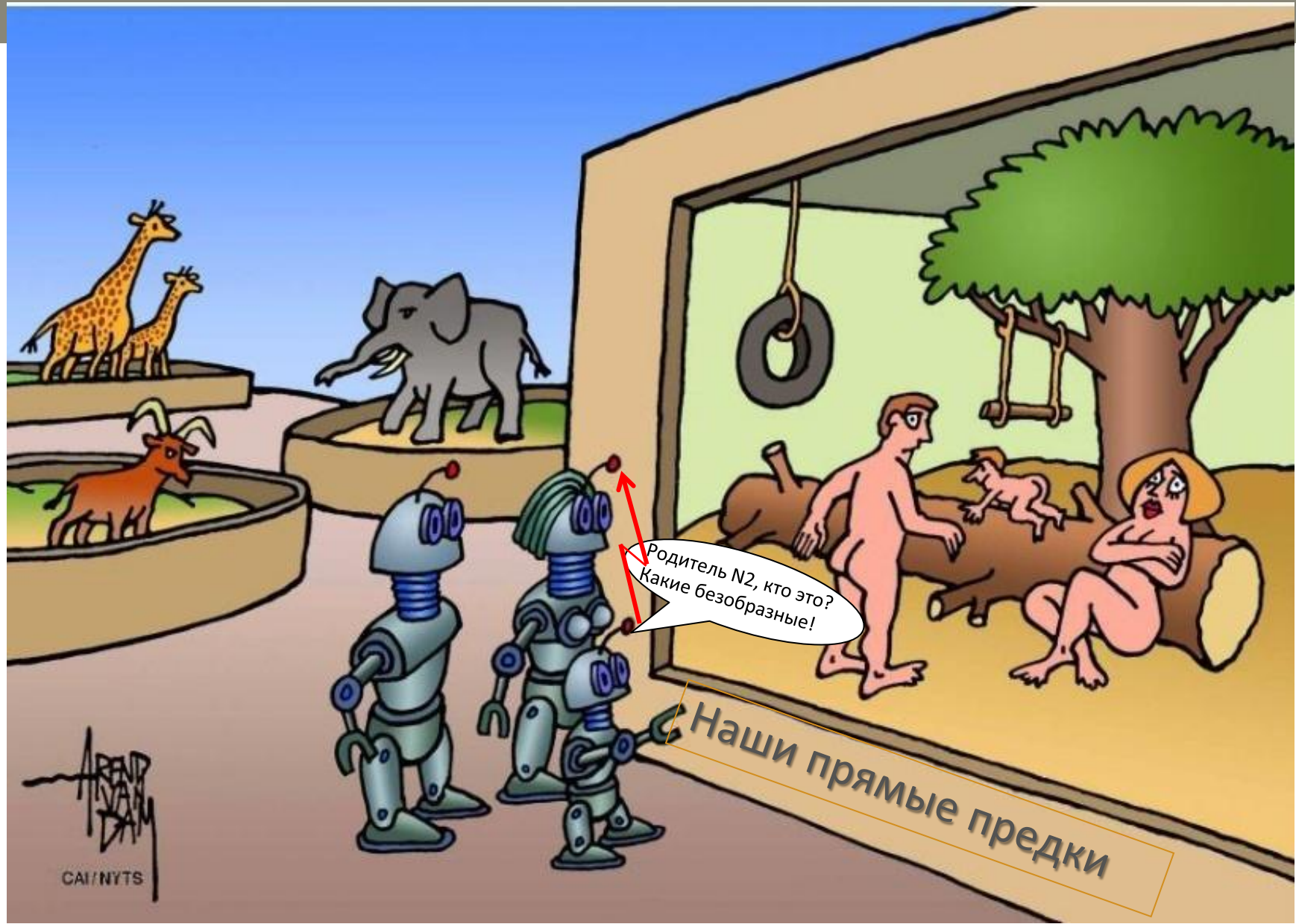
Ю.Г.Карпов

Синтез супервизоров



В области построения контроллеров для координации действий распределенных дискретных 'событийных' систем получены интересные результаты

2100 год. Мы не успели ...



Родитель N2, кто это?
Какие безобразные!

Наши прямые предки

CAI/NYTS

Заключение.



- Роботы, супермозг и интеллектуальные

технологии - это сложные программные

комплексы, «поведение» которых сложно контролировать

- Как реализовать системы ИИ так, чтобы они были **ГАРАНТИРОВАННО** контролируруемыми, управляемыми и human-friendly?- важнейшая задача компьютерных наук
- Формализация и проверка **ТРЕБОВАНИЙ** к поведению интеллектуальных систем возможно с помощью **модальной логики**. **ОТВЕТСТВЕННОСТЬ ПРОГРАММИСТОВ** при создании мульти-агентных систем **ОГРОМНА**.
- Впереди у специалистов в области компьютерных наук **увлекательный тяжелый** путь **творческой** практической работы и научных исследований.

Итак: Потенциально возможно то, что можно вычислить

- **Принцип «познание через *моделирование*».** Познание традиционно осуществляется посредством моделирования свойств реальности в сознании субъекта, то есть выполнения операций и анализа их последствий не с самими объектами реальности, а с символами-понятиями или носителями «познанных» свойств объектов.
- **В 20 веке** неотъемлемым компонентом познания стало **компьютерное моделирование** - описание объектов реальности с помощью программ количественной характеристики потенциально возможных свойств объектов и систем. Компьютерное моделирование породило т.н. виртуальную реальность – «цифровую тень» физической реальности.

Так может выглядеть «потенциально возможное»



Горизонт событий – то, что разделяет «возможное» и «состоявшееся»



Моделью «возможного и состоявшегося» является река, вода в которой ускоряет свой бег на пути к водопаду. Существует граница, которая **разделяет ламинарный и турбулентный потоки**.

В макром мире роль водопада играет, например, «черная дыра». То место и тот момент, где/когда скорость потока вещества превышает скорость распространения гравитационных волн, называется **горизонтом событий**. Можно ли построить компьютерную модель того, что происходит за этим горизонтом ?

Системные свойства: эмерджентность и ингерентность

Эмерджентность - феномен, который наделяет систему свойствами, которые не присущи ни одной из ее составных частей

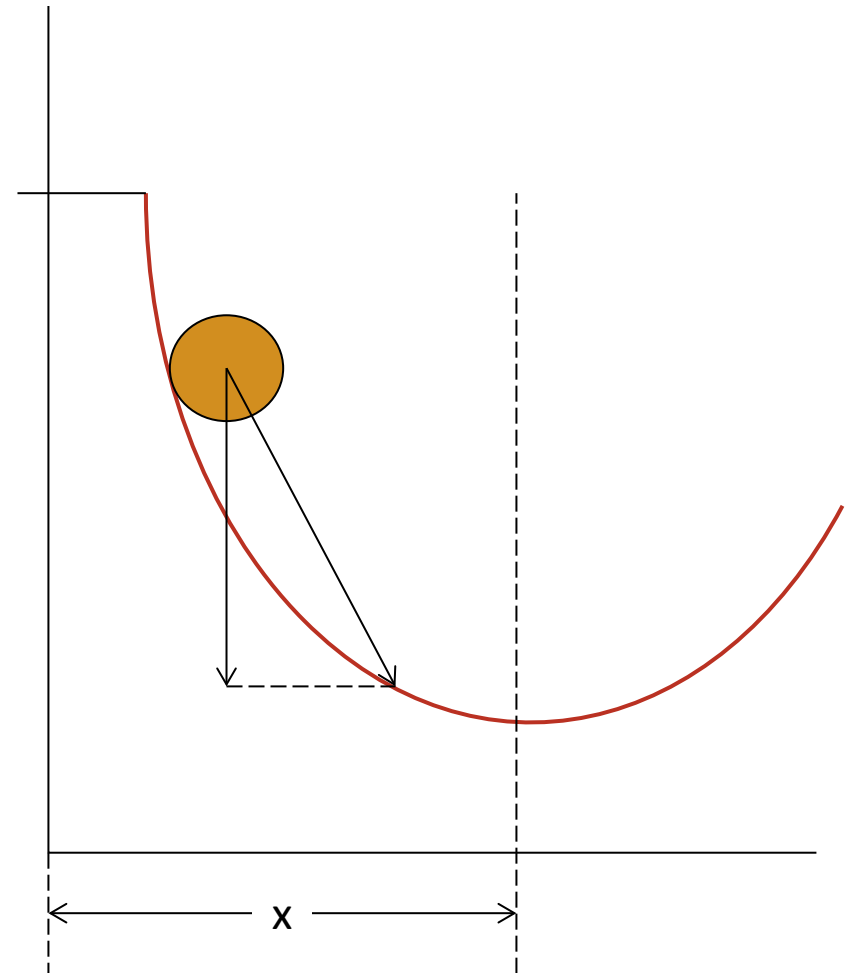
Ингерентность - способность системы выполнять заданную функцию в определенной окружающей среде (от англ. inherent -являющийся неотъемлемой частью чего-то).

Физико-механическая «теория мира»

Онтологические сущности физическо-механической реальности - **силы**, которые порождают движение пространственно-временных структур.

Следствия:

- Все свойства системы могут быть получены как арифметическая «сумма» свойств составляющих ее компонент (с учетом их размерности)
- Все возможные изменения структур и процессов подчиняются законам, сформулированным по отношению к «идеальным объектам»
- «Наблюдатель» происходящих процессов находится вне наблюдаемой системы



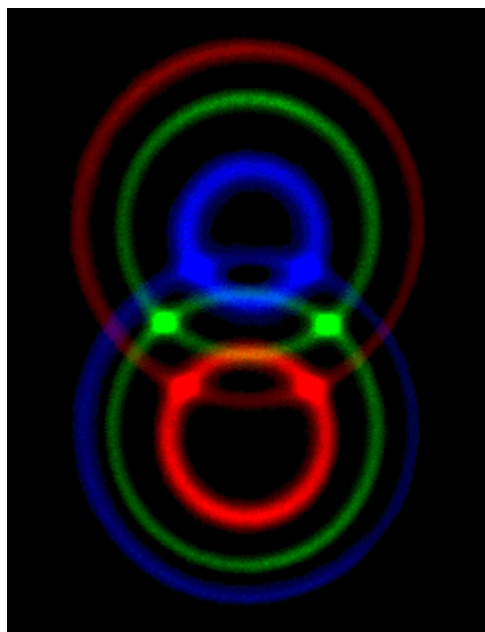
Информационно-вычислительная «теория мира»

- Онтологические сущности И-В реальности: **вычисления**, которые порождают изменения информации (структуры) материи
- Системы наделяются эмерджентными свойствами, которые НЕ могут быть получены исключительно из свойств составляющих систему компонент
- Изменения материальных структур подчиняются законам «открытых систем», ингерентность которых есть следствие самоорганизации и адаптации.
- «Наблюдатель» процессов есть составная часть наблюдаемой системы

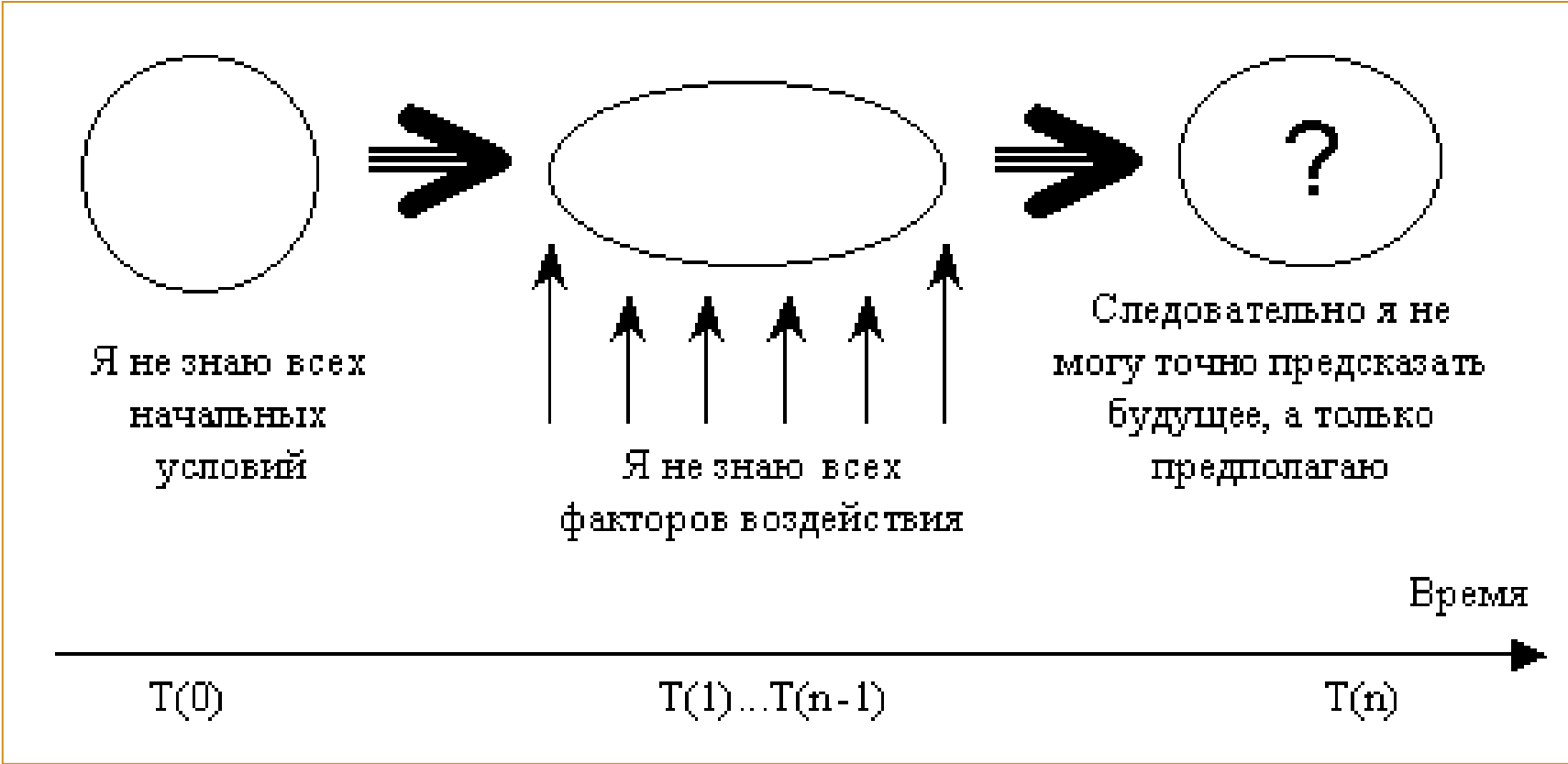


Рефлексия реальности

Невозможно вывести свойства реальности исходя только из аддитивного объединения свойств отдельных подсистем.



Концепция возможного vs состоявшегося



Потенциальная возможность : классическая или квантовая

Классические (неквантовые) представления о возможном исходе исходят из того, что случайность является «ненастоящей» (субъективной).



На самом деле объект в любой момент времени обладает определенными значением параметра и до и после измерения, только до измерения это значение «скрыто от нас», а измерение просто проявляет то, что было ранее скрыто (кубик имел определенное «состояние» и до того как его вынули из урны). Хотя в урне кубики существуют «сепарабельно», но некоторые их свойства «спутаны», например, суммарное число граней или средний «выигрыш».

Спутанность проявляется не в физическом (функциональном гильбертовом), а в вероятностном или информационном пространстве.

Потенциально дискретное, квантовое и непрерывное

Для пространственных координат и его импульса (произведение массы на скорость) действует соотношение неопределенности:

$$\Delta X \Delta P \geq \frac{\hbar}{2}$$

Судя по всему, на фундаментальном уровне материя (вещество и поле) «дискретны»:

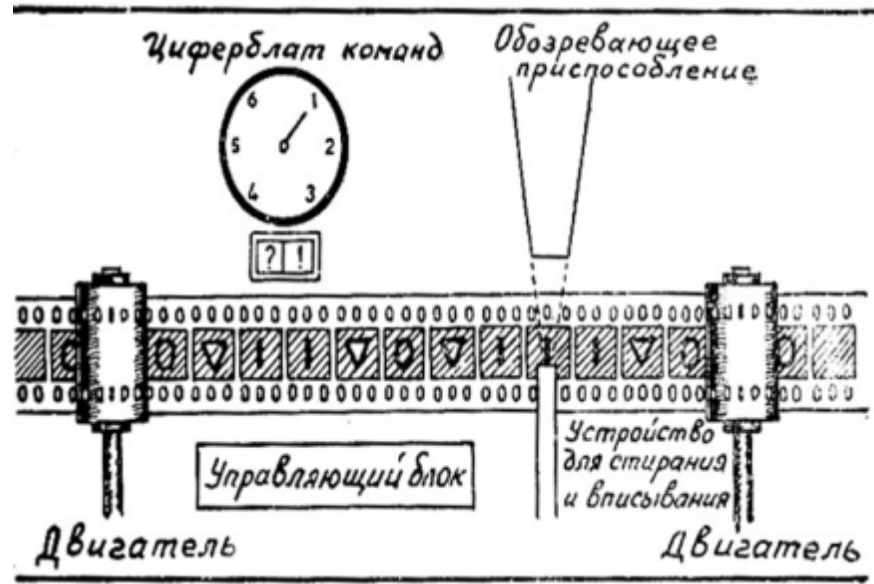


Постоянная Планка указывает нижний точности **знания** физических величин, после которого для описания физических процессов надо учитывать из «дискретность» или квантовые свойства, где $\hbar = h / 2\pi = (1,0545887 \pm 0,000007) * 10^{-27}$ эрг * с



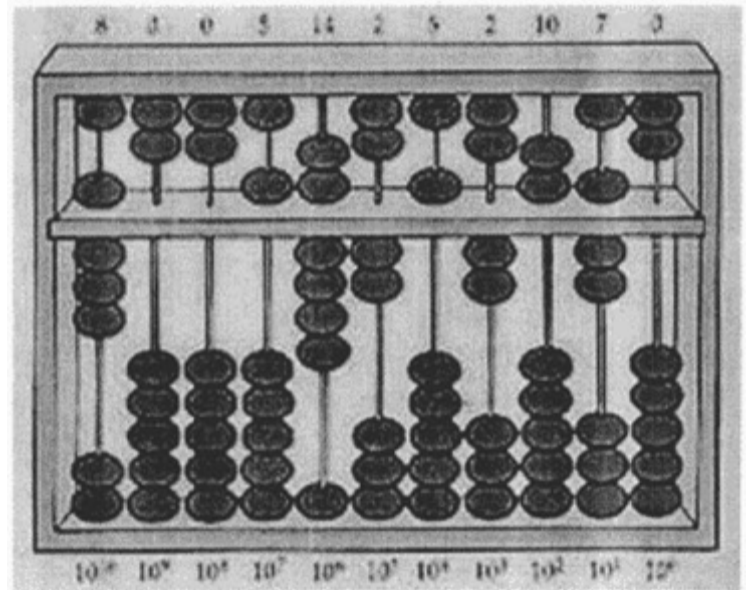
... а «все что мы знаем о реальности «сделано» из составляющих, которые не могут считаться реальными» (Н. Бор)

Вычисления устраняющие неопределенность - машина Тьюринга (МТ)



МТ - это автомат с «бесконечной цифровой лентой», движение которой есть «вычисление» нового символа, который кодирует полученный результат.

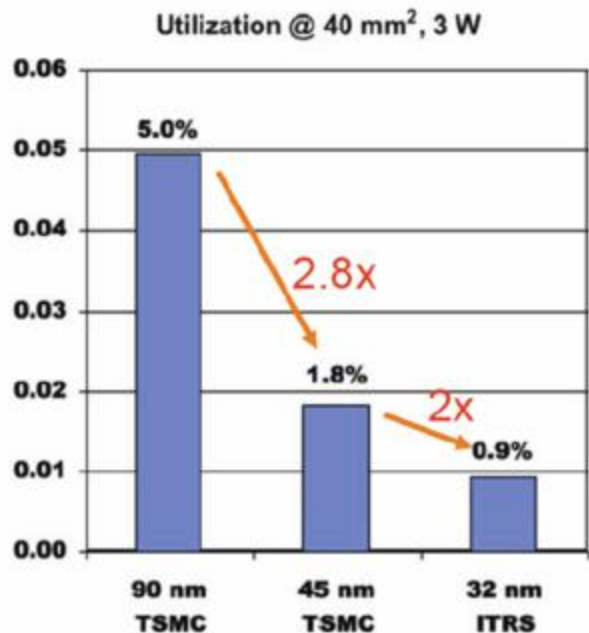
Физическое «поведение» МТ – основано на работе двигателей, которые осуществляют «перемещение ленты», на которой расположены символы «алфавита», подсчете количества перемещений, устройства «обозревающего» результат и запоминания, полученных результатов.



Системные ограничения процессов вычислений

- Закон Мура и теория Деннарда: при масштабном уменьшении параметров вычислительных процессов в S раз, «объемная» вычислительная мощность (в идеальных условиях) возрастает в S^3 .
- При сохранении удельных энергетических затрат на вычисления **доля площади кристалла микропроцессора, задействованной в активной работе** (область кристалла, где могут переключаться транзисторы) **убывает экспоненциально**.
- В современных условиях «активная» кристалла микропроцессора площадь измеряется единицами или даже долями процентов от всей площади. Оставшаяся большая часть кристалла, не задействованная в данный момент в работе, получила название **«тёмный» кремний** («Dark» Silicon).

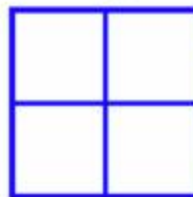
«Стена утилизации»



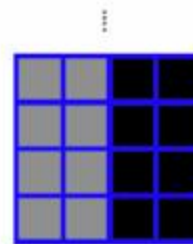
Spectrum of tradeoffs
between # of cores and
frequency

Example:
65 nm → 32 nm (S = 2)

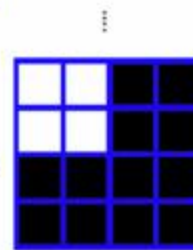
4 cores @ 1.8 GHz



65 nm



2x4 cores @ 1.8 GHz
(8 cores dark, 8 dim)
(Industry's Choice)

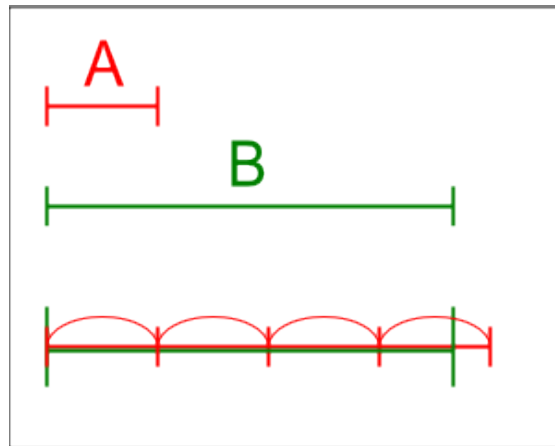


4 cores @ 2x1.8 GHz
(12 cores dark)
*75% dark after 2 generations;
93% dark after 4 generations*

В многоядерных микропроцессорах в любой момент работы большая часть кристалла является «темным» кремнием – бездействует или работает на значительно пониженной тактовой частоте.

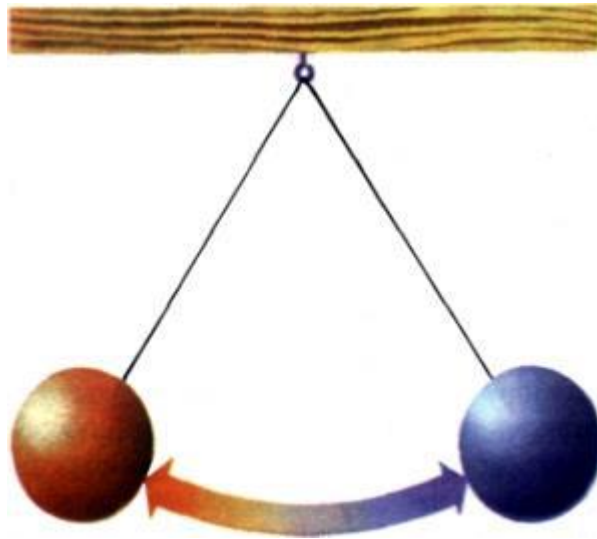
Аксиома Архимеда «классической вычислительной математика»

Аксиомой Архимеда называется такое утверждение: если даны отрезки A (масштаб) и B (объект измерения) , то можно так отложить отрезок A несколько раз, что сумма будет равна или «немного» превосходить отрезок B



Утверждение : изучаемое математикой ПРОСТРАНСТВО-ВРЕМЯ архимедово то есть **одномасштабно**, значит гладко, «делимо и однородно».

«Простые» физические объекты – одномасштабны

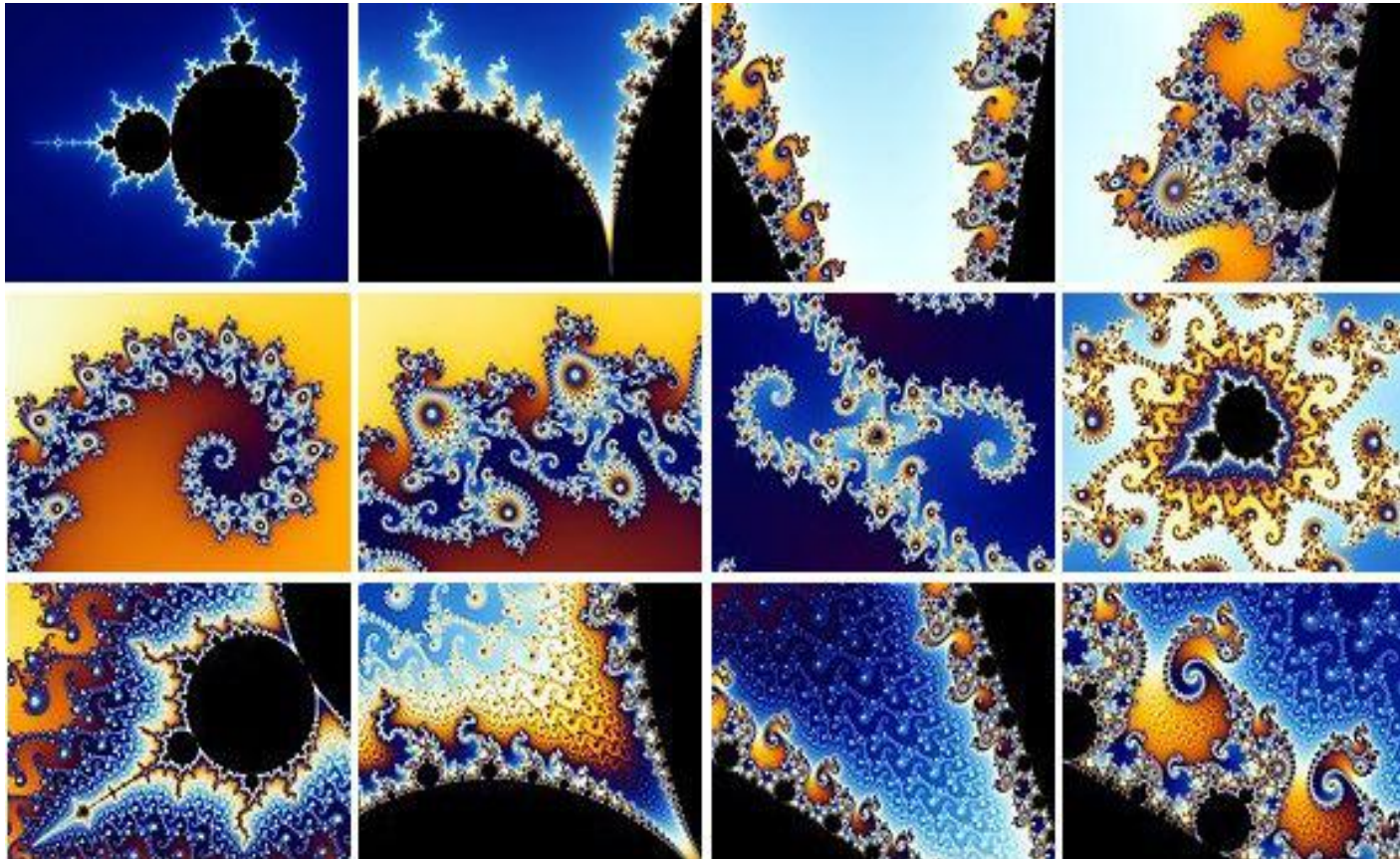


«Сложные» объекты классической физики мультимасштабны, но ...их состояния не спутаны

Сложные структуры
не однородны

они не имеют одного
масштаба, поэтому
связаны

это значит, что «деление»
на «подобъекты»
невозможно



Закон исключенного третьего «исключает» спутанность состояний макрообъекта

В «реальном» мире суперпозиции нет - происходит коллапс «вектора состояний» или редукция «волновой функции» к одной из возможностей

Пример:
«Кот Шредингера»:



Ключевой вопрос: можно ли «остановить» редукцию волновой функции и, если да, то на какое время?

Итого

- Основная идея «математики возможных состояний» в том, что вычисления с квантовыми объектами – кубитами, позволяют использовать информацию о состоянии этого объекта передать эту информацию из одного места в другое за конечное время
- Всем прочитать **S. Lomonaco**, “A Rosetta Stone for Quantum Computation”