

**Перспективы искусственного
интеллекта и теорема Пенроуза.
Часть I. Грозит ли нам
технологическая сингулярность.**

А. Д. Панов
НИИЯФ МГУ

Мотивация

- Проблема SETI и модели будущего
- Два рукава эволюции. И третий рукав?

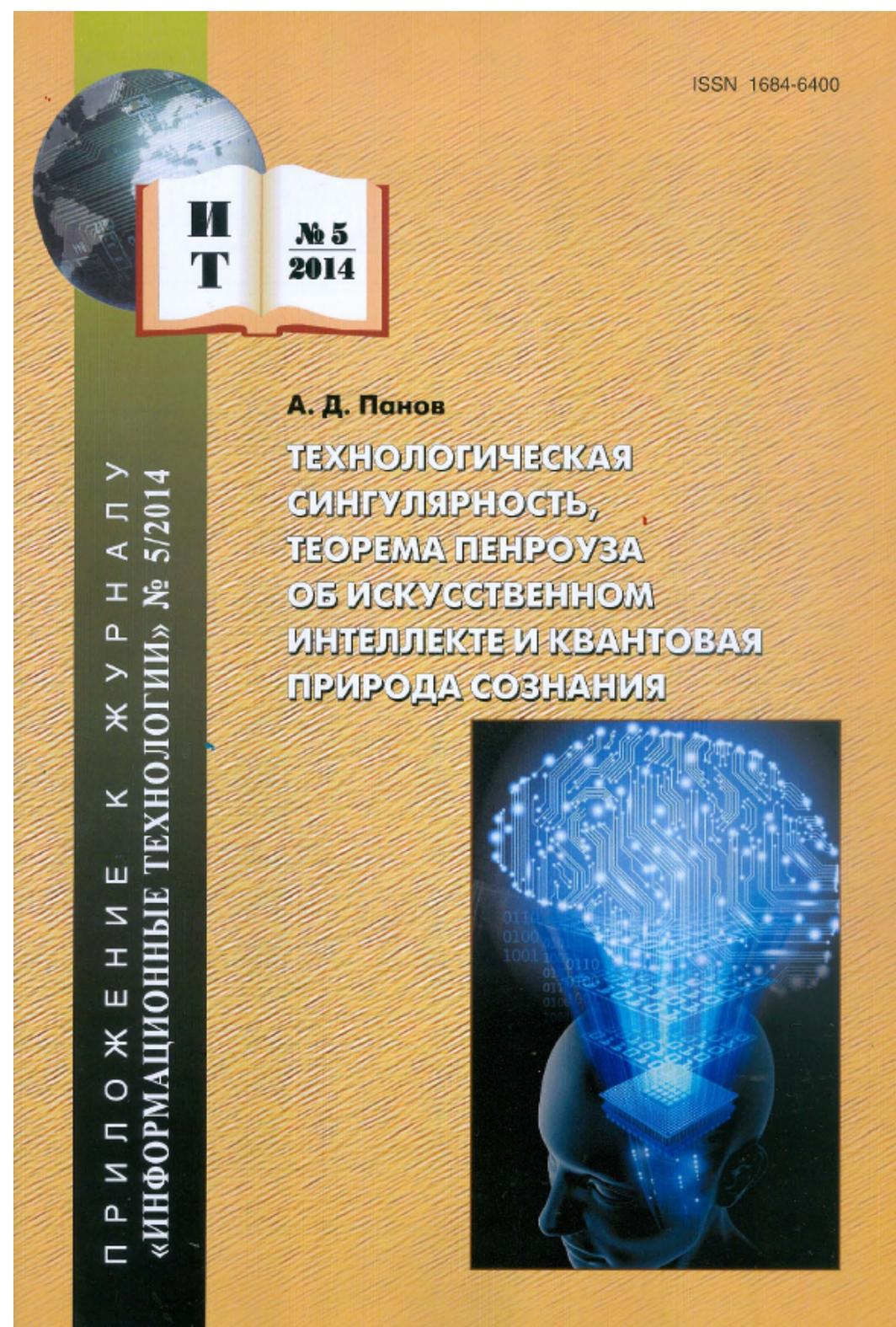
Вывод:

В настоящее время нет никаких указаний на возможность создания сильного ИИ в любом обозримом будущем



Художник - Александр Пинкин

Первый раз
(приблизительно) этот
доклад был сделан в
2012 г.



Искусственный Интеллект -

автономное (искусственное) устройство,
способное к интеллектуальной деятельности.

Инструментальный интеллект -

человеческий интеллект,
использующий инструментальные средства:
калькулятор, компьютер,
электронные базы данных,
бумагу и карандаш, книги...

Четкой границы между искусственным и инструментальным интеллектом нет.

Технологическая сингулярность

Искусственный Интеллект (ИИ)
превосходит человеческий

во всех отношениях

(сильный ИИ)

→ предсказание будущего
становится невозможным

WIRED, April 2000

Why the future doesn't need us.

Our most powerful 21st-century technologies
– robotics, genetic engineering, and nanotech –
are threatening to make humans an endangered species.

By Bill Joy

From the moment I became involved in the creation of new technologies, their ethical dimensions have concerned me, but it was only in the autumn of 1998 that I became anxiously aware of how great are the dangers facing us in the 21st century. I can date the onset of my unease to the day I met Ray Kurzweil, the deservedly famous inventor of the first reading machine for the blind and many other amazing things.

Ray and I were both speakers at George Gilder's Telecosm conference, and I encountered him by chance in the bar of the hotel after both our sessions were over. I was sitting with John Searle, a Berkeley philosopher who studies consciousness. While we were talking, Ray approached and a conversation began, the subject of which haunts me to this day.

I had missed Ray's talk and the subsequent panel that Ray and John had been on, and they now picked right up where they'd left off, with Ray saying that the rate of improvement of technology was going to accelerate and that we were going to become robots or fuse with

robots or something like that, and John countering that this couldn't happen, because the robots couldn't be conscious.

While I had heard such talk before, I had always felt sentient robots were in the realm of science fiction. But now, from someone I respected, I was hearing a strong argument that they were a near-term possibility. I was taken aback, especially given Ray's proven ability to imagine and create the future. I already knew that new technologies like genetic engineering and nanotechnology were giving us the power to remake the world, but a realistic and imminent scenario for intelligent robots surprised me.

It's easy to get jaded about such breakthroughs. We hear in the news almost every day of some kind of technological or scientific advance. Yet this was no ordinary prediction. In the hotel bar, Ray gave me a partial preprint of his then-forthcoming book *The Age of Spiritual Machines*, which outlined a utopia he foresaw – one in which humans gained near immortality by becoming one with robotic technology. On reading

Что такое человеческий интеллект (разум)?

То, что отличает людей даже от самых умных животных, что позволило людям создать технологическую цивилизацию, и чего не хватило для этого другим живым существам.

Не свойство каждого отдельного человеческого мозга, но лишь свойство коллективов людей.

Что значит превзойти человеческий интеллект (во всех отношениях)?

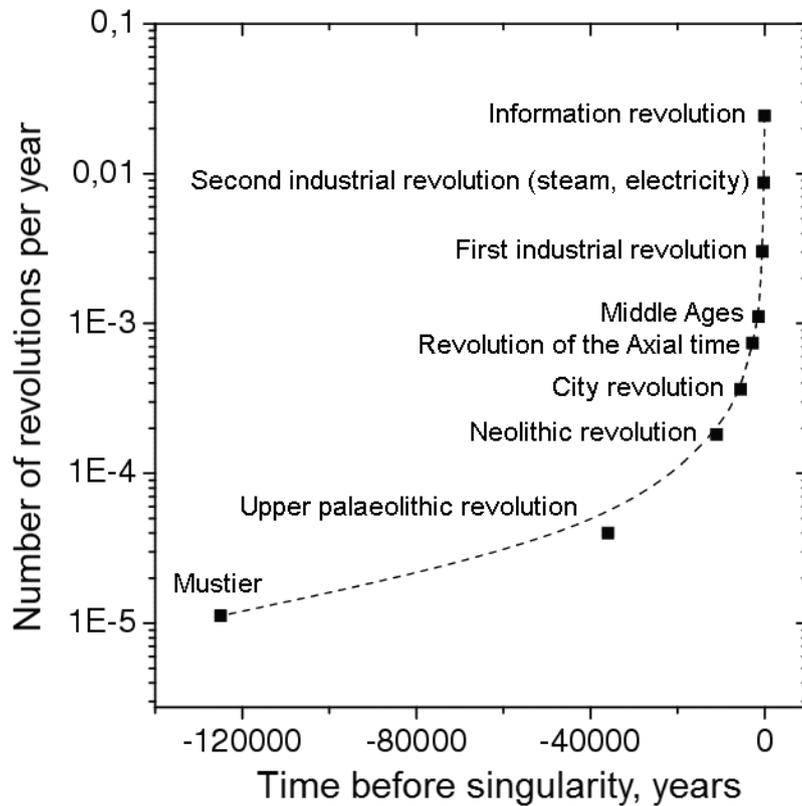
Это значит — превзойти коллективный разум всего человечества **в наивысших проявлениях**

А по каким параметрам компьютеры еще не превзошли человеческий разум?

- Универсальная способность к пониманию
- ≈ Способность создавать новое понимание
- Свобода воли

Не путать технологическую сингулярность с различными эволюционными сингулярностями!

Общая эволюционная сингулярность.



$t^* \sim 2000-2050$

Демографическая сингулярность

$$N = \frac{c}{t^* - t}; \quad t \rightarrow t^* \Rightarrow N \rightarrow \infty$$

$$1/N = (t^* - t)/c$$



И.С. Шкловский,
1965

$t^* \sim 2030$

История понятия и основатели

Irving John Good, 1965 -

intelligence explosion

Vernor Vinge, 1988 -

technological singularity,

2005-2030

Hans Moravec, 1988 -

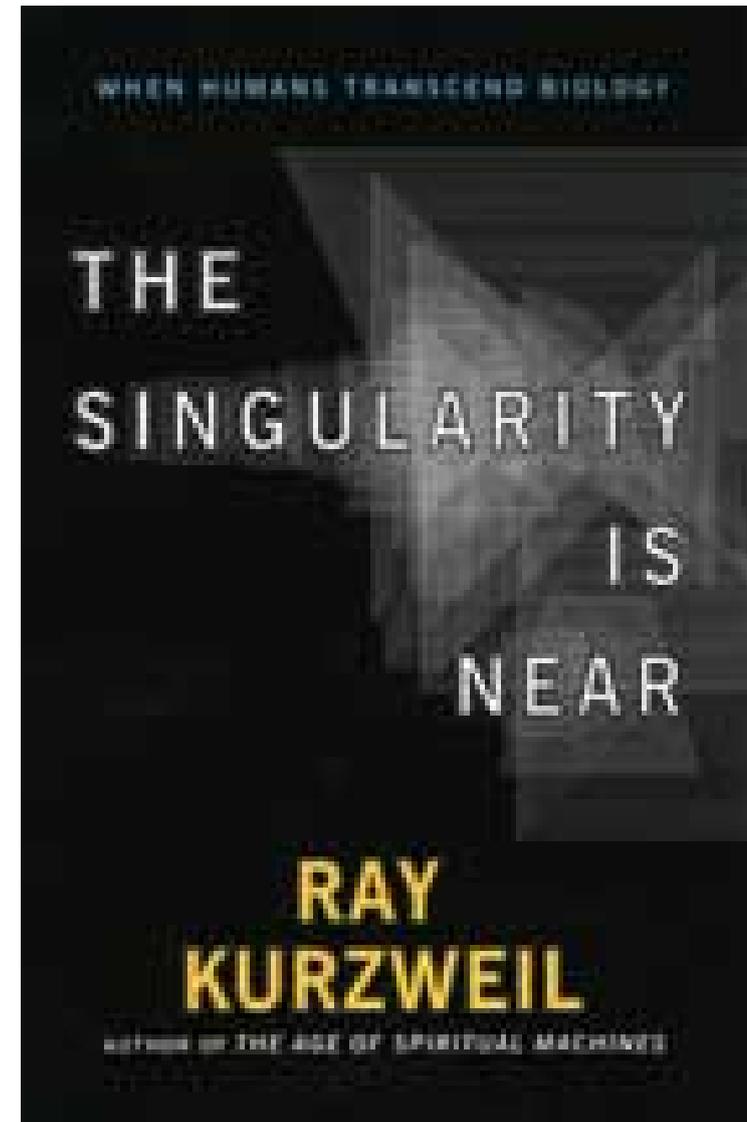
technological singularity,

2030-2040

Ray Kurzweil, 1990th -

technological singularity,

2045



2005

Когда? Рэй Курцвейл:

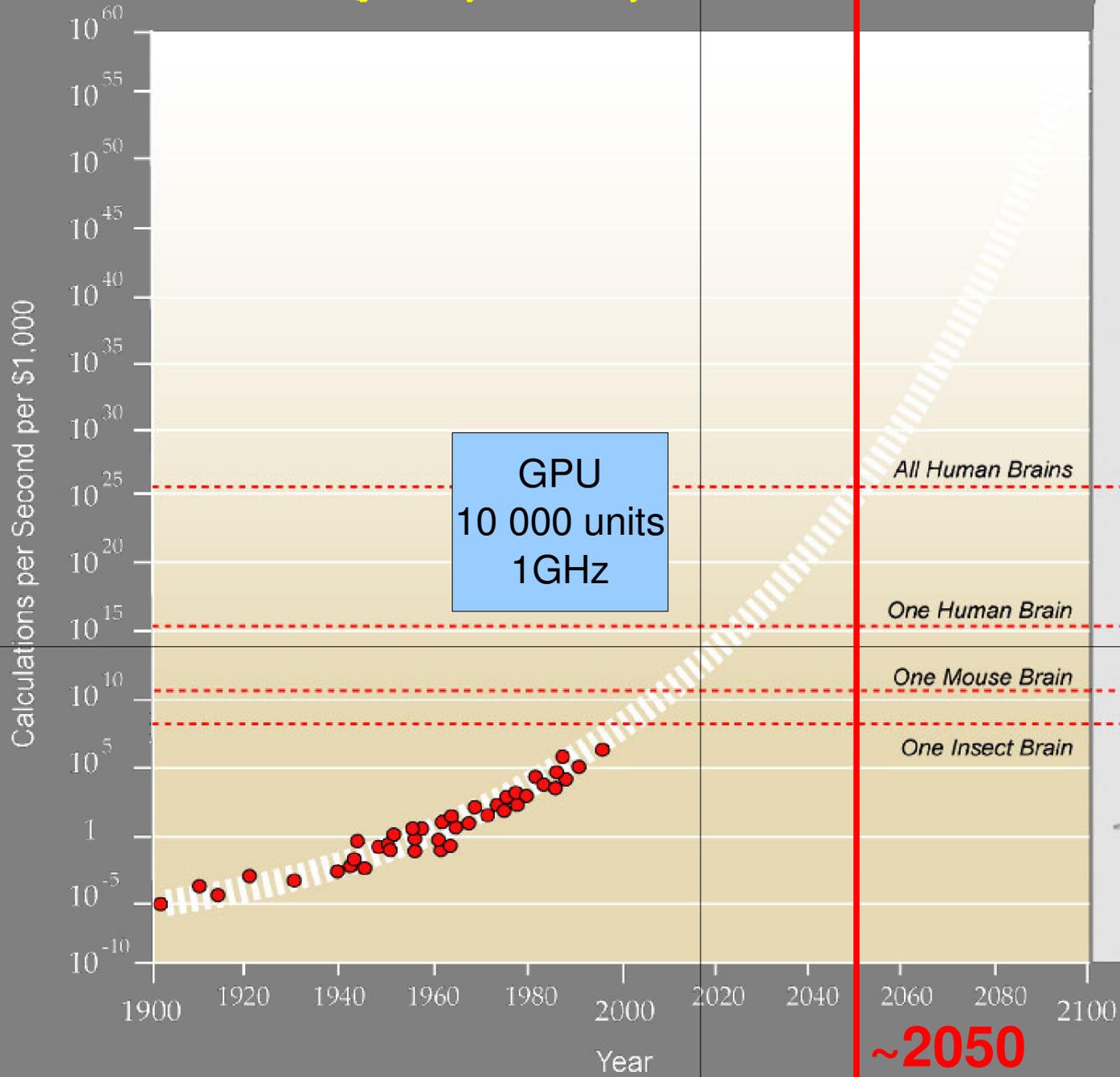
Как только мощность коммерческих компьютеров, выраженная в операциях в секунду, превзойдет совокупную вычислительную мощность мозга всего человечества, сильный ИИ будет создан, и технологическая сингулярность может стать реальностью.

Вопросы:

1. Как растет скорость компьютеров?
2. Какова вычислительная мощность мозга?

Exponential Growth of Computing

Twentieth through twenty first century



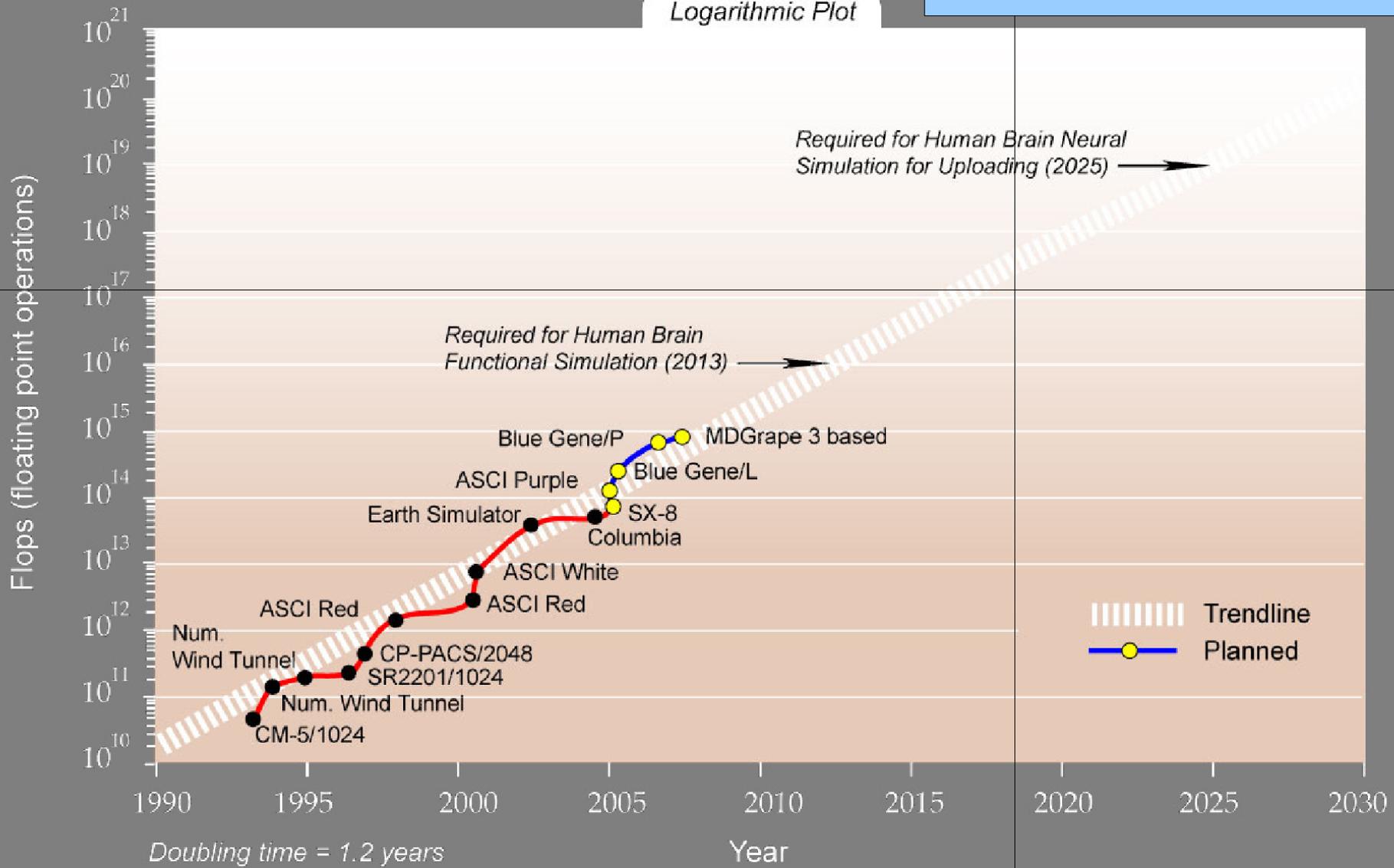
Logarithmic Plot



кластер Summit от IBM
 4356 серверов
 по два 22-ядерных CPU
 по шесть GPU NVIDIA Tesla V100
 122 петафлопс

Growth in
Supercomputer Power

Logarithmic Plot



Основные направления критики

Три плохо обоснованных предположения

1. Переоценка фактора мощности компьютера и недооценка фактора программного обеспечения в создании ИИ
2. Возможная недооценка скорости вычислений мозга
3. Аналогия мозг-классический компьютер

Одно полностью не понятое обстоятельство

No-go теорема Пенроуза об ИИ -

Компьютеры классической архитектуры не могут привести к созданию сильного ИИ в принципе

Переоценка фактора мощности компьютера и недооценка фактора программного обеспечения в создании сильного ИИ

Как только мощность компьютеров превысит
совокупную мощность мозга всего человечества
то возникновение сильного ИИ будет обеспечено.

За прошедшие 15 лет «разум» наших электронных вычислительных машин улучшился в миллион раз... В течение нескольких следующих десятилетий следует ожидать увеличения характеристик «разума» машин еще по крайней мере в несколько десятков тысяч раз.

«Разум» таких машин по основным параметрам будет

заведомо превосходить разум человека.

За прошедшие 15 лет «разум» наших электронных вычислительных машин улучшился в миллион раз... В течение нескольких следующих десятилетий следует ожидать увеличения характеристик «разума» машин еще по крайней мере в несколько десятков тысяч раз. «Разум» таких машин по основным параметрам будет **заведомо** превосходить разум человека.

Иосиф Самуилович Шкловский, 1975, Земля и Вселенная

43 года прошло!

Компьютерные мощности увеличились примерно **в 10 миллиардов раз** (с 10^7 флоп до 10^{17} флоп).

Где же машины, которые заведомо превосходят разум человека по основным параметрам?

Понимание ситуации было и остается фундаментально неверным.

Где корень ошибки в предсказании?

Для того, чтобы создать сильный ИИ,
мало иметь достаточные компьютерные мощности.

Надо знать, КАК создать сильный ИИ.

Нужны соответствующие методы, нужно
программное обеспечение, нужно понимание того,
какую задачу решать.

Программное обеспечение гораздо более
консервативно, чем аппаратное обеспечение
(и пользовательские интерфейсы).

Два основных пути в создании ПО для ИИ:

1. Синтетическое направление

2. Обратная инженерия мозга

Синтетическое направление создания ИИ

maxima — одна из лучших современных систем аналитических вычислений.

Mathematica, Maple имеют то же ИИ ядро.

Классическая система ИИ (эвристическое программирование).

Была создана в 1972 году, 46 лет назад.

Компьютерные мощности возросли в ~10 миллиардов раз.

Microsoft Word — версия для Windows была создана в 1989 г., 29 лет назад.

Компьютерные мощности возросли в 10^6 - 10^7 раз.

Где же интеллектуальные роботы - секретари?

Компьютерный перевод — качество почти не выросло с начала 1990-х,

Компьютерные мощности возросли в 10^6 - 10^7 раз.

«Робот-учёный делает открытия без помощи человека»?

<http://www.infuture.ru/article/1917>

«Адам сформулировал и протестировал 20 гипотез относительно генетических кодов 13 ферментов»

Проблема четырёх красок и т. д.

Направление активно развивается

Искусственная жизнь — Адаптивное поведение (animat)

Направление активно развивается, но результаты хоть и очень интересные, но пока довольно скромные.

Основные технологии программирования ИИ

Нейронные сети

1958, Ф. Розенблатт, однослойный перцептрон.

1974, Пол Дж. Вербос и А. И. Галушкин - алгоритм обратного распространения ошибки для обучения многослойных перцептронов.

1982. J.J. Hopfield - сети Хопфилда, спиновые стекла.

Эвристическое программирование

1957, The General Problem Solver (GPS) Newell, Shaw and Simon

Экспертные системы

1965, DENDRAL — расшифровка данных масс-спектрологии

1974, PROSPECTOR — обнаружение полезных ископаемых

1975, MYCIN — диагностика инфекционных заболеваний

Эволюционное (генетическое) программирование

Начало 1960-х, J.H. Holland

1973, I. Rechenberg

1977, H.-P. Schwefel

Обучение с подкреплением, deep learning

Возможно, этих технологий принципиально не хватает.

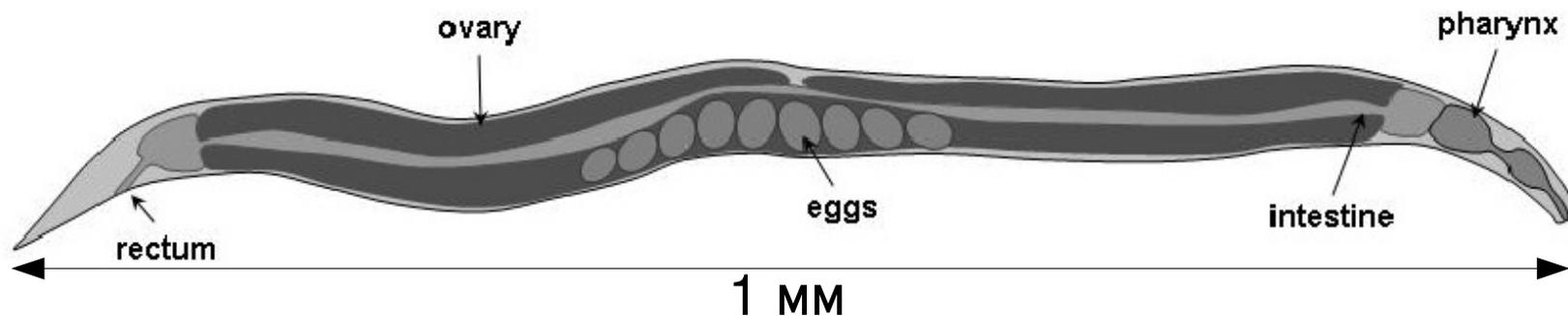
Прогноз не ведет к созданию сильного ИИ в обозримом будущем.

Экстраполяция нуля в любое будущее есть ноль.

Обратная инженерия мозга:

Достаточные компьютерные мощности
не гарантируют решения задачи.

Caenorhabditis elegans (C. Elegans) - нематода



- Нервная система взрослой особи — гермафродита состоит из **302 нейронов**
Проблема мощности компьютера полностью отсутствует!
- Сложный репертуар поведений:
навигация, поиск пищи, спаривание, обучение, социальное поведение, сон.
- Нейронная система полностью картирована, каждый нейрон и синапс имеют свое имя.
- С середины 1990-х делаются попытки симулировать нервную систему *C. Elegans* на компьютере. <http://www.jefftk.com/news/2011-11-02>
- Результат очень ограничен. Только симуляция движений тела (есть статьи).
- **Проблемы:** сила синаптических связей, **тело, среда обитания.**

А. Г. СВЕРДЛИК

КАК ЭМОЦИИ
ВЛИЯЮТ НА
АБСТРАКТНОЕ МЫШЛЕНИЕ
И ПОЧЕМУ
МАТЕМАТИКА
НЕВЕРОЯТНО
ТОЧНА



КАК УСТРОЕНА КОРА ГОЛОВНОГО МОЗГА,
ПОЧЕМУ ЕЁ ВОЗМОЖНОСТИ ОГРАНИЧЕНЫ
И КАК ЭМОЦИИ, ДОПОЛНЯЯ РАБОТУ КОРЫ,
ПОЗВОЛЯЮТ ЧЕЛОВЕКУ
СОВЕРШАТЬ НАУЧНЫЕ ОТКРЫТИЯ



«Внутренняя ось» - тело - принимает
существенное участие в мышлении

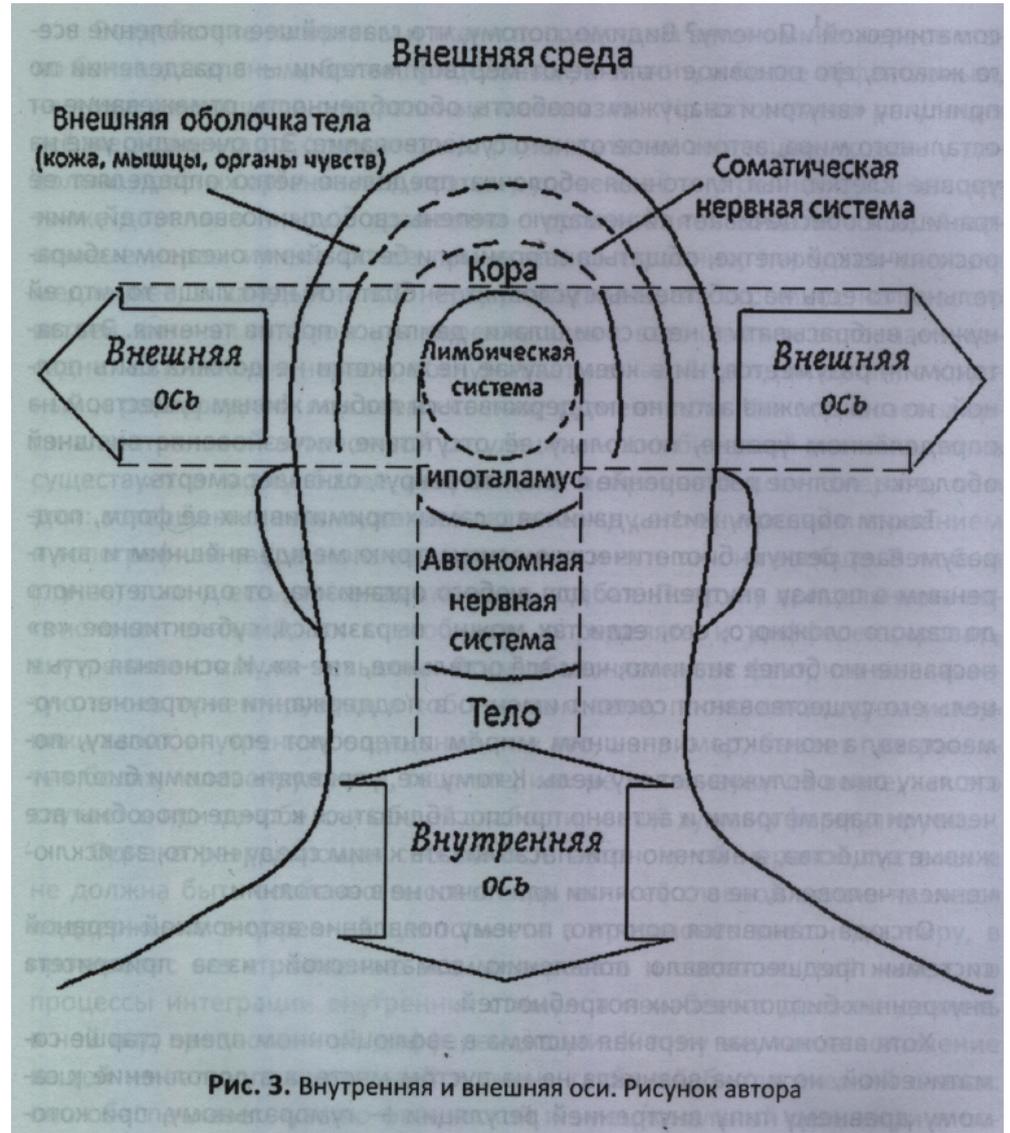


Рис. 3. Внутренняя и внешняя оси. Рисунок автора

**Мыслит не мозг,
мыслит Вселенная в целом.**

**Мозг есть лишь современное
оконечное устройство вселенского мышления
(front-end device)**

**Невозможно провести границу между
материальными агентами и явлениями,
которые имеют отношение к процессу мышления,
и которые не имеют отношения к процессу мышления.**

**Звездное небо над нами и нравственный закон
внутри нас не только наполняют душу благоговением,
но и неразрывно связаны в едином процессе
мышления.**

- **Чтобы смоделировать мышление человека в компьютере, нужно в нем смоделировать не мозг, но всю Вселенную.**
- **Либо нужно создать полноценного «робота», связанного миллионами связей с людьми, другими роботами, всей Вселенной.**
- **Либо робот будет не антропоморфен, либо его интеллект не будет сильнее человеческого.**
- **Не антропомарфный ИИ есть еще одно окончное устройство для мышления Вселенной. Вопрос о том, кто кого превосходит снимается.**

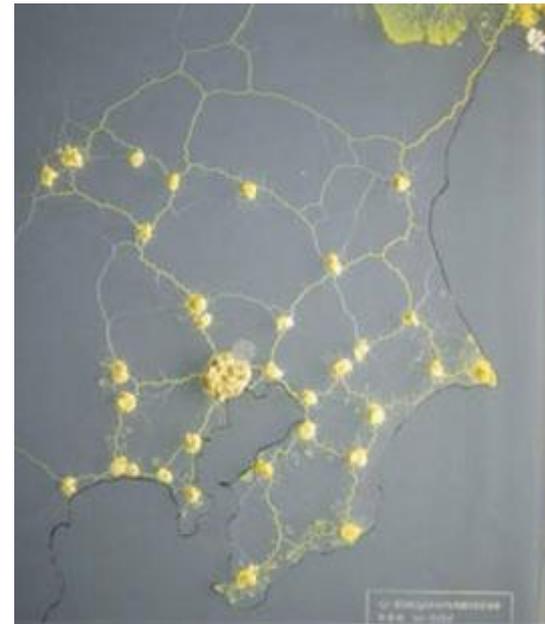
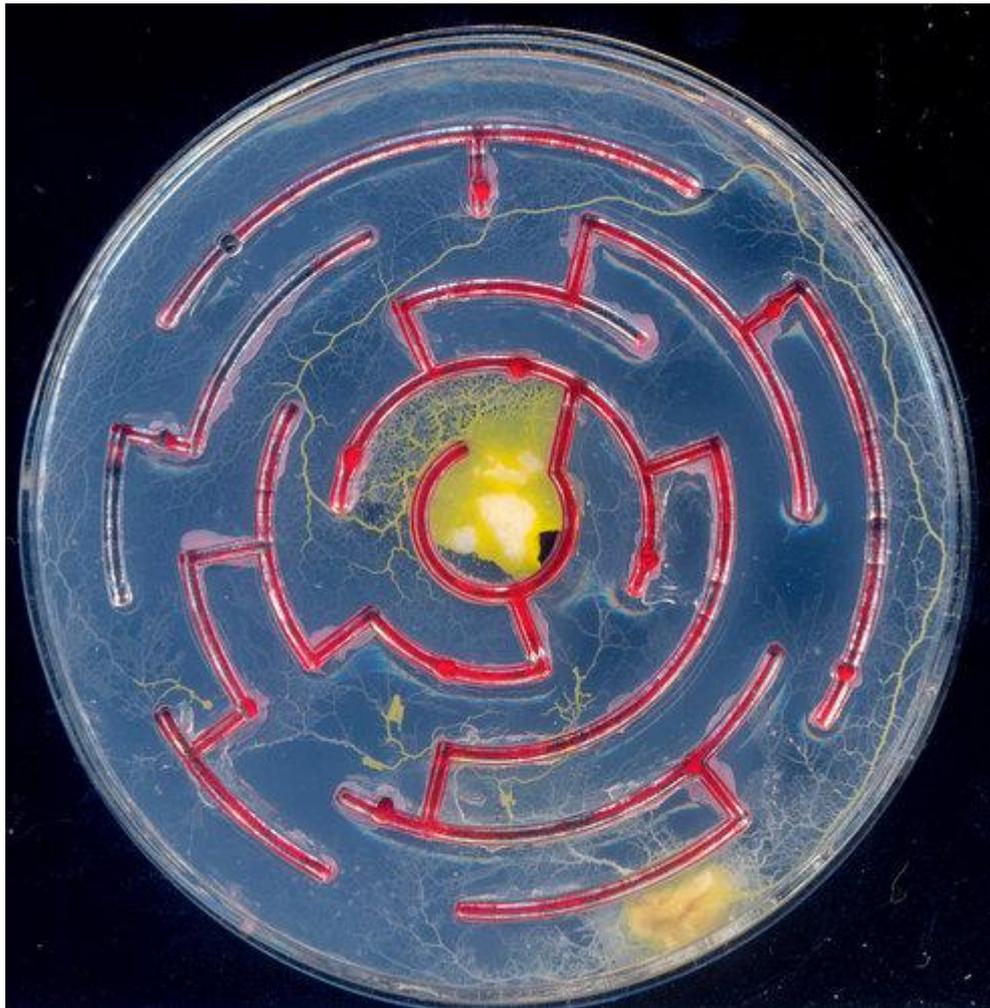
Проблема не только и не столько в том, что не хватает компьютерной мощности, сколько в том, что непонятно, как решать задачу (и даже — какую задачу надо решать).

**Возможная недооценка
скорости вычислений мозга.**

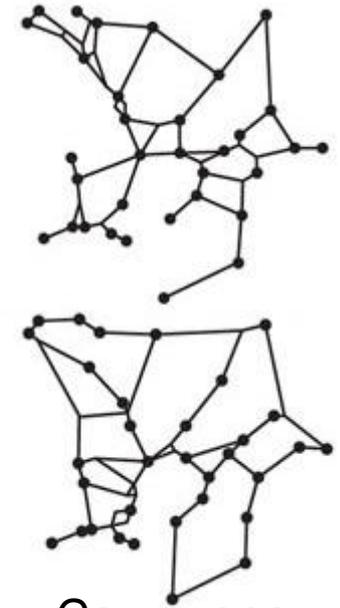
К быстродействию мозга

**Является ли только
нейросетевая активность
«носителем» сознания?**

Слизевик: лабиринты, оптимизация, обучение.



Сеть слизевика



Сеть дорог

Влажный воздух заставляет миксомицетов двигаться быстрее, а сухой – наоборот, замедляет перемещение. Чередую поток влажного и сухого воздуха, ученые со временем зафиксировали интересную особенность: перед очередной подачей сухого воздуха слизевика снижали скорость.

<http://lostlab.ru/forum/topic432.html>

Эволюционный консерватизм.

Ничто, созданное эволюцией, не пропадает зря.

Механизмы «мышления» одноклеточных должны работать и в нейронах, в том числе мозга

УДК 62-506+612.815.1

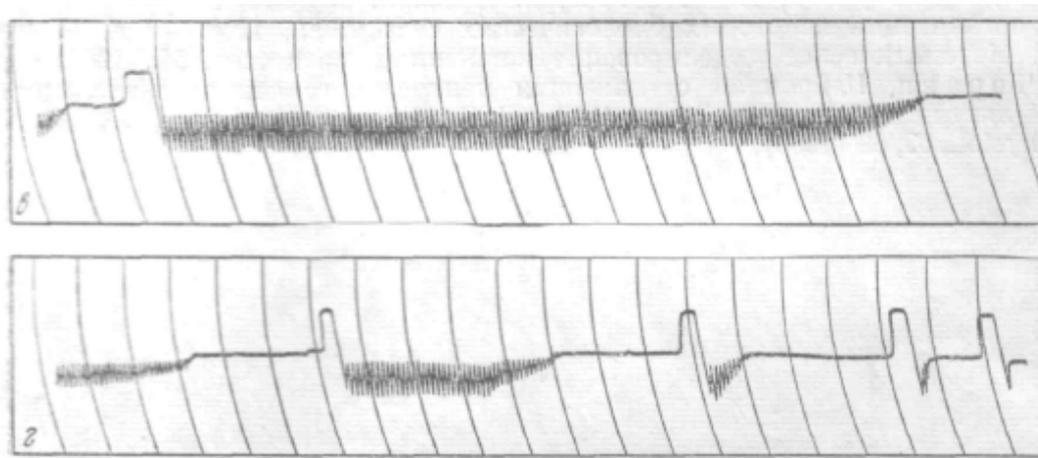
ФИЗИОЛОГИЯ

Б. Г. РЕЖАБЕК

О ПОВЕДЕНИИ МЕХАНОРЕЦЕПТОРНОГО НЕЙРОНА
В УСЛОВИЯХ ЗАМЫКАНИЯ ЕГО ЦЕПЬЮ ИСКУССТВЕННОЙ
ОБРАТНОЙ СВЯЗИ

(Представлено академиком Е. М. Крепсом 21 IV 1970)

Нейрорецептор растяжения речного рака — не имеет синаптических связей с другими нейронами, не входит в нейронную сеть.



Обучение индивидуального нейрона.
Сохраняется несколько часов.

Где на субклеточном уровне может
помещаться механизм
«внутриклеточного сознания»?

Цитоскелет и микротрубочки

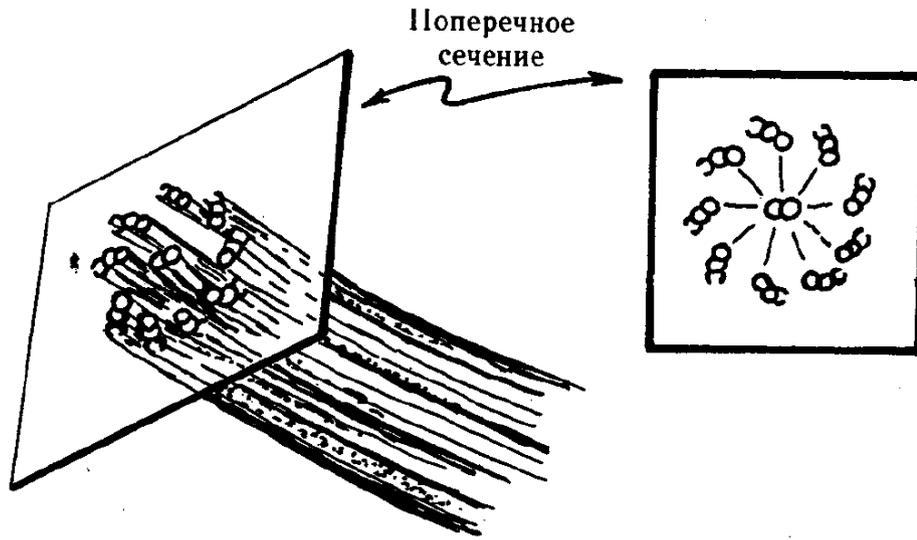


Рис. 7.3. Важной частью цитоскелета являются пучки крохотных трубочек (микротрубочек), организованных в структуры, напоминающие в поперечном сечении лопасти вентилятора. Такое строение имеют, например, реснички парамеции.

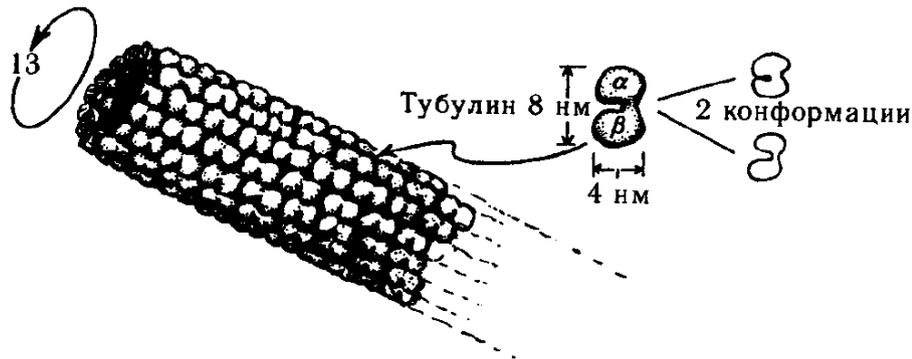


Рис. 7.4. *Микротрубочка*. Полая трубка, обычно состоящая из 13 рядов димеров тубулина. Каждая из молекул тубулина может существовать в двух (по крайней мере) конформациях.

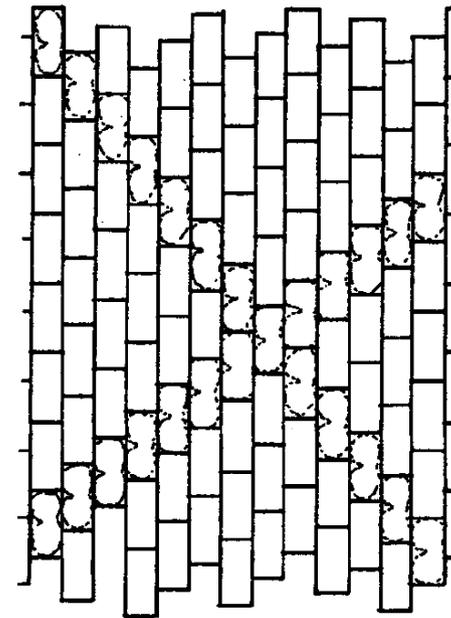
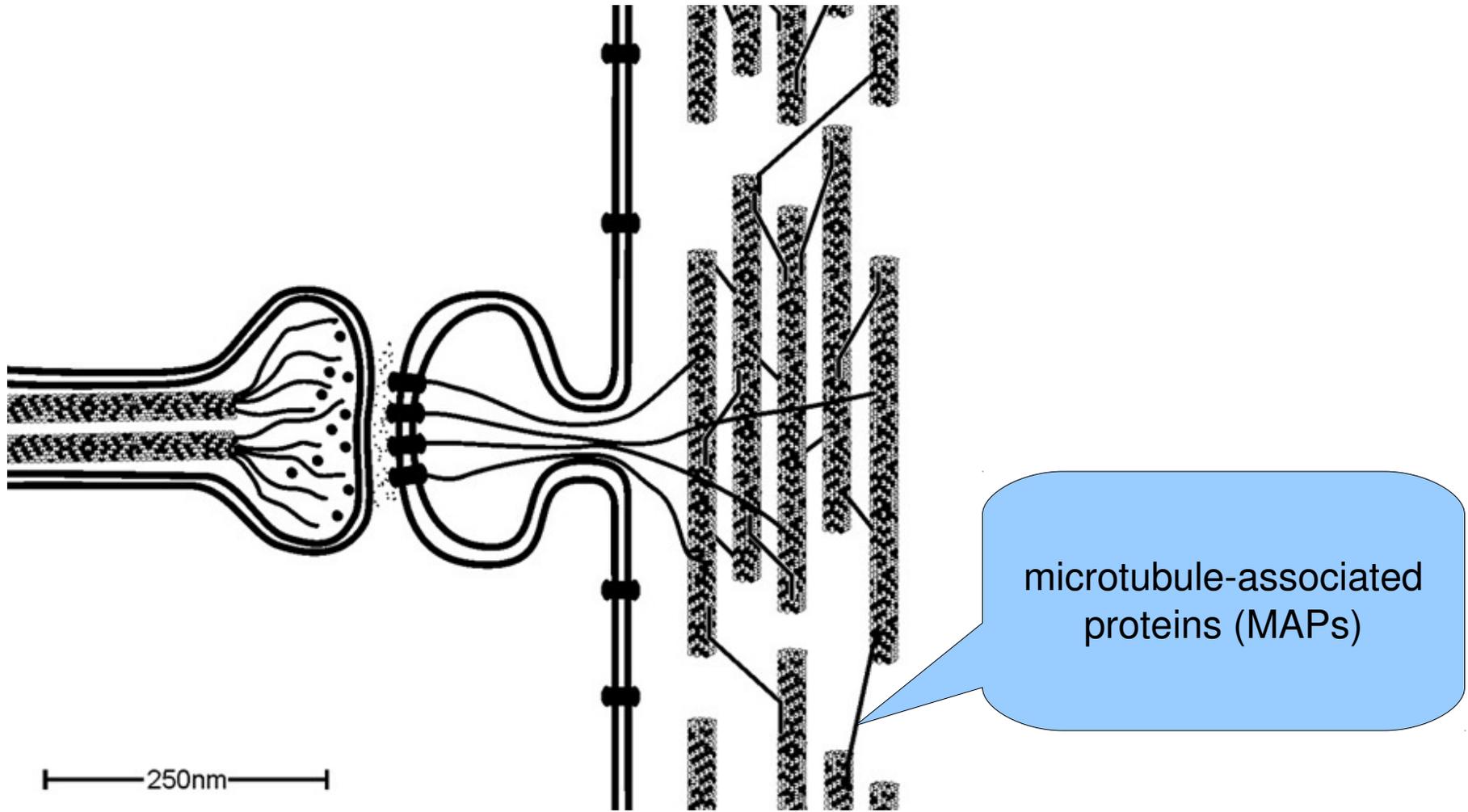
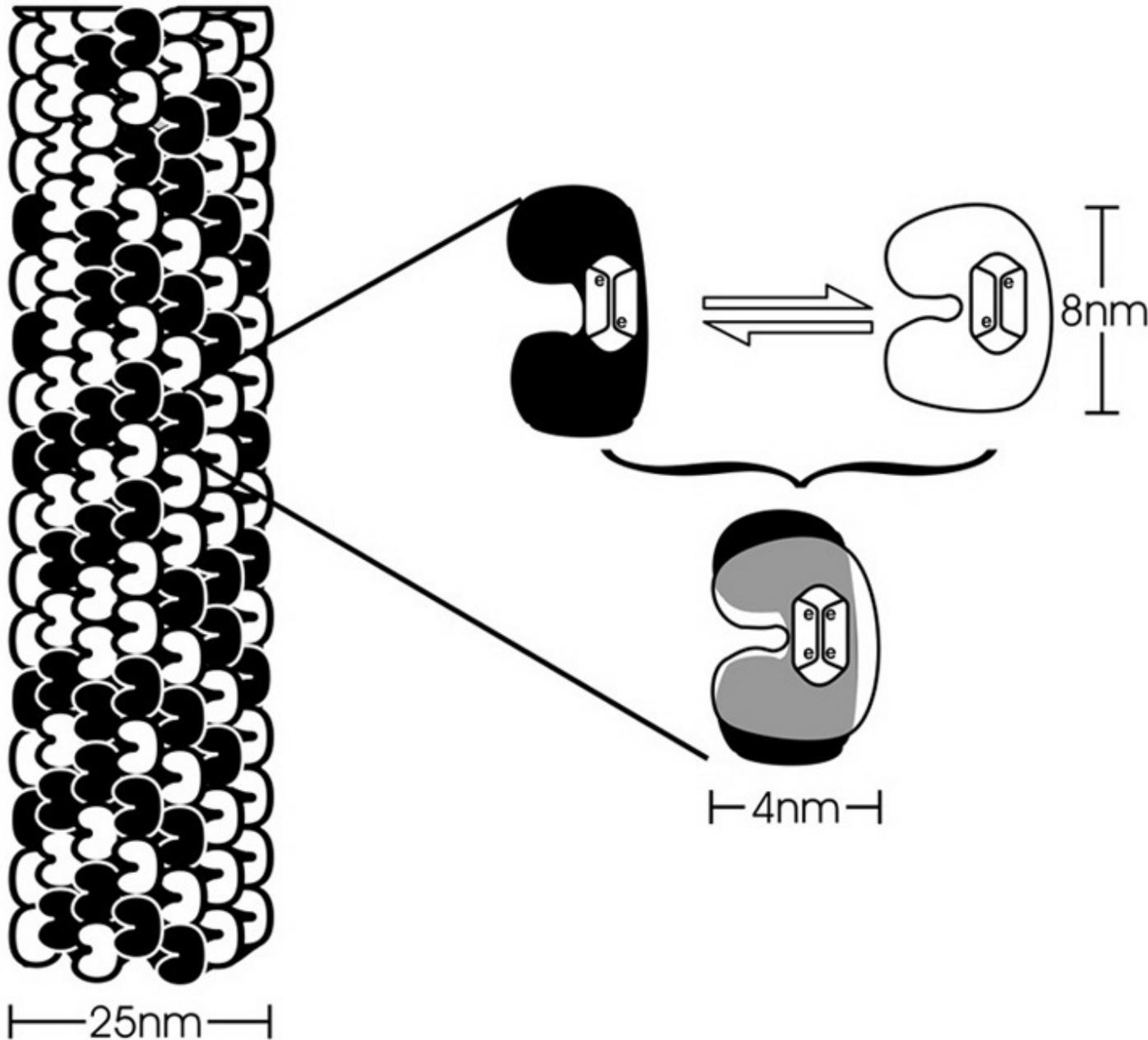


Рис. 7.8. Представим, что микротрубочка разрезана вдоль и затем развернута в полосу. Можно видеть, что молекулы тубулина располагаются вдоль наклонных линий, причем каждый новый виток смещен относительно предыдущего на 5 или 8 молекул (в зависимости от того, куда наклонена линия, вправо или влево).

Микротрубочки объединены в сети



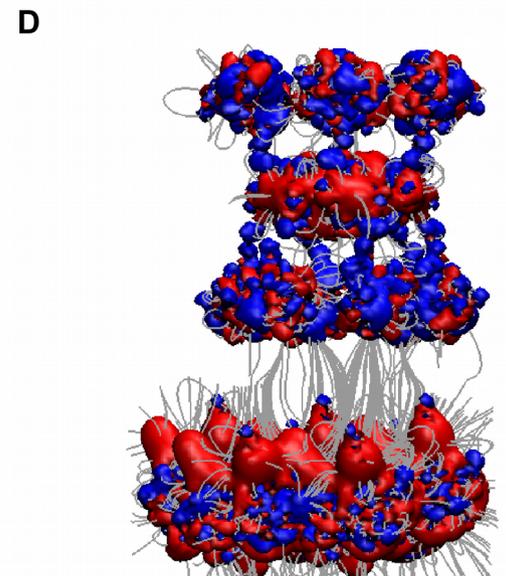
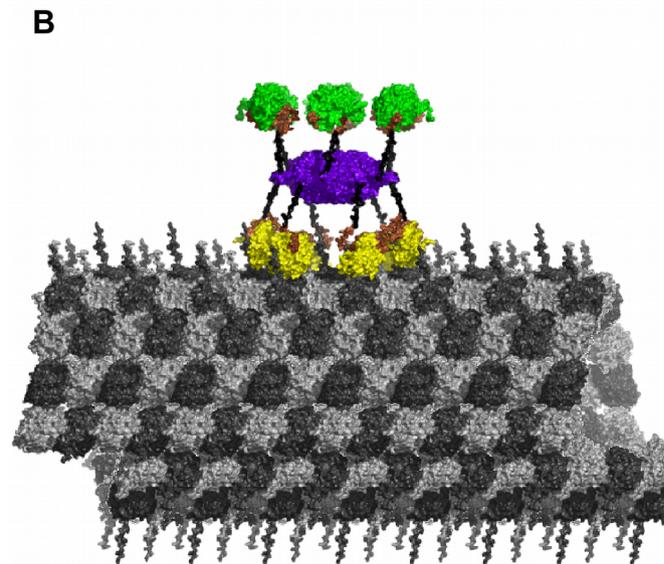
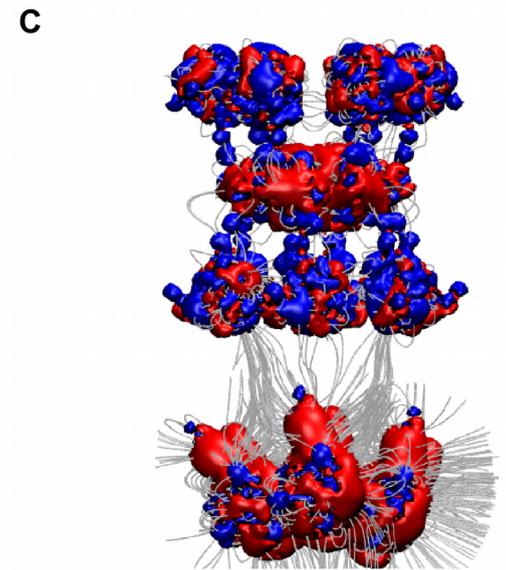
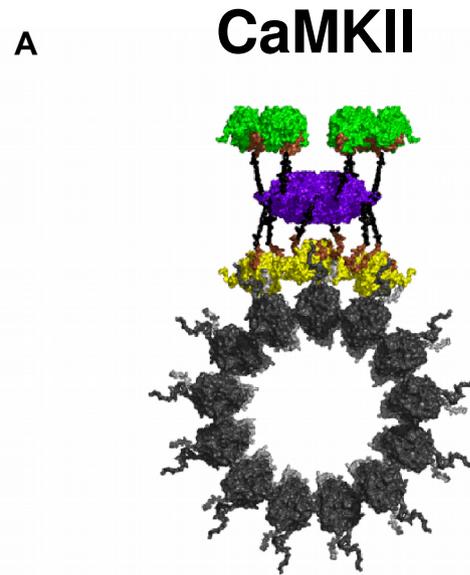
Микротрубочка — клеточный автомат, состоящий из молекул тубулина — битов с двумя состояниями



α - тубулин
 β - тубулин
две
конформации,
0/1

CaMKII holoenzyme (протеинкиназа II) - 6-битная кодировка «памяти» микротубочек фосфорилированием тубулина

Craddock TJA,
Tuszynski JA,
Hameroff S (2012)
Cytoskeletal Signaling:
Is Memory Encoded in
Microtubule Lattices by
CaMKII Phosphorylation?
PLoS Comput Biol 8(3):
e1002421.
doi:10.1371/journal.pcbi.1002421



CaMKII играет также ключевую роль в
регуляции синаптической пластичности!

Каково быстроедействие «клеточного автомата» микротрубки?

10^7 Гц — основная частота микротубочек:

Pokorny, J., Hasek, J., Jelnek, F., Saroch, J. & Palan, B. (2001)
Electromagnetic activity of yeast cells in the M phase.
Electro Magnetobiol 20, 371–396.

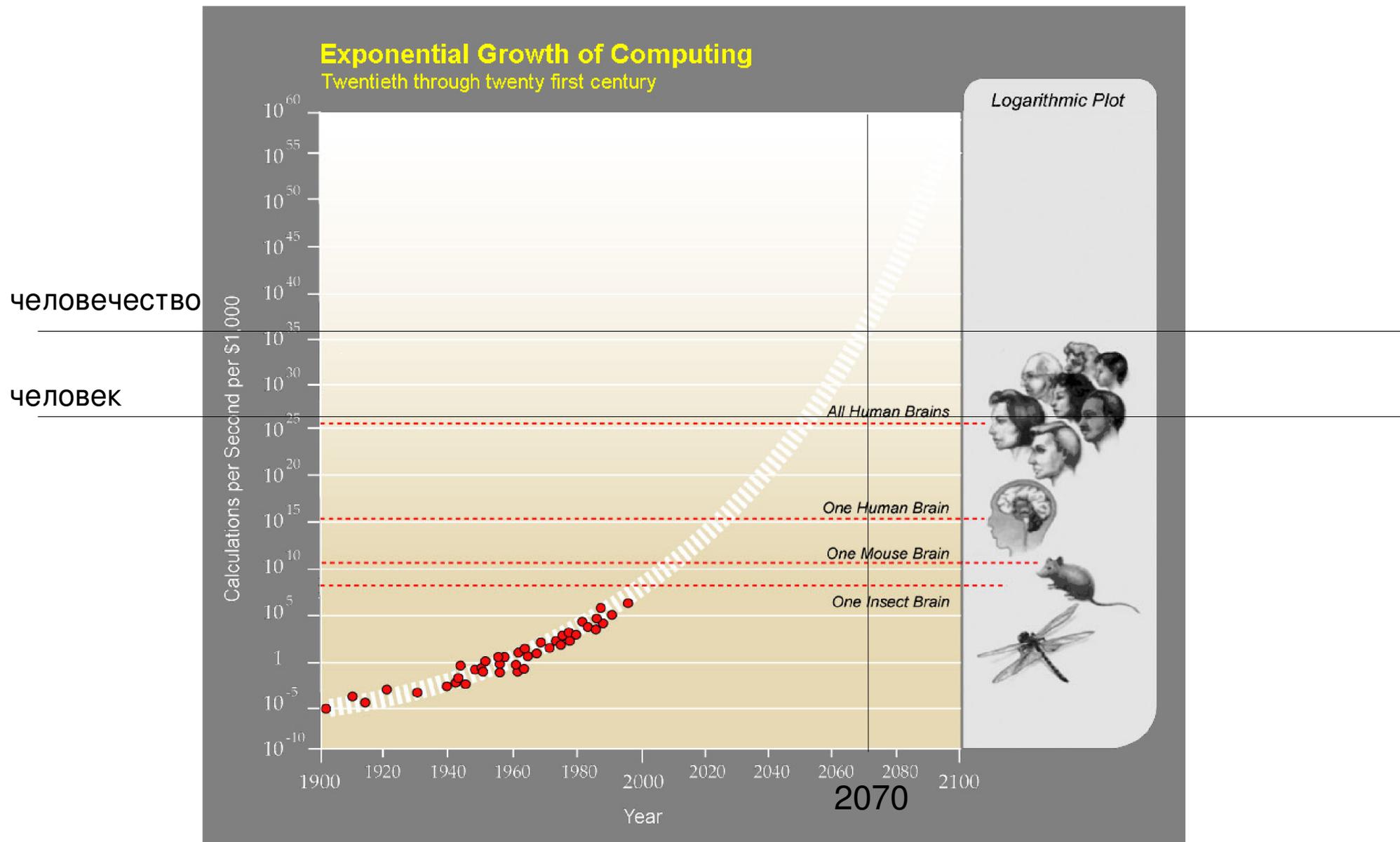
Pokorny, J. (2004)
Excitation of vibration in microtubules in living cells.
Bioelectrochem. 63: 321-326.

Каково быстродействие «клеточного автомата» микротрубки?

На один нейрон: 10^8 тубулинов $\times 10^7$ Гц = 10^{15} Гц

На мозг: 10^{11} нейронов $\times 10^{15}$ Гц = 10^{26} Гц

На 10 порядков
больше



Аналогия МОЗГ-классический компьютер

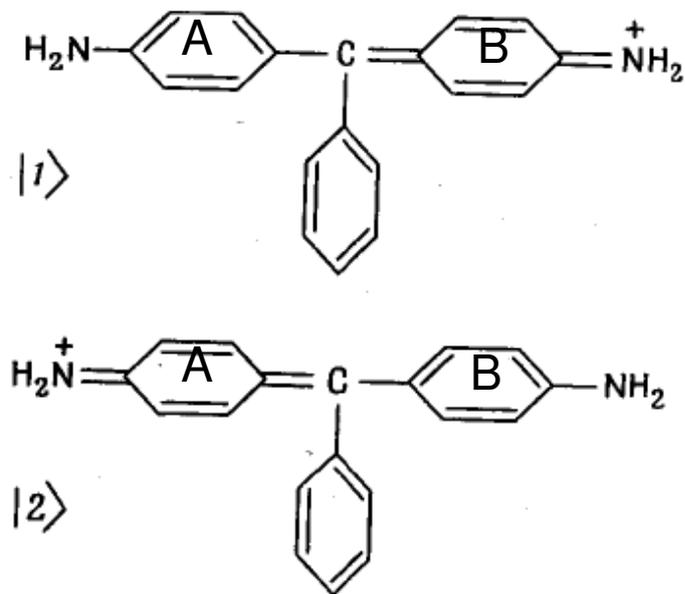
Аналогия мозг — это классический компьютер.

Является ли мозг классическим вычислительным устройством?

- Может ли мозг быть квантовым компьютером?
- Быстрые процессы нейронной сети —
(скорее всего) классические процессы.
- Могут ли на более глубоких уровнях работать квантовые процессы вычислений?

- Спинтроника и сверхнизкие температуры -
наивный контр-аргумент (Рэй Курцвейл).
- Атом гелия. Температура ионизации 25000 градусов.
- Долгоживущие нестационарные состояния.
Люминофоры. Десятки часов.

Квантовые суперпозиции в макромолекулах



Фуксин, краситель.
 При комнатной температуре —
 кристалл или раствор.

$$|I\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle); \quad E$$

$$|II\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle); \quad E + \Delta E$$

$$|1\rangle = |A3\rangle|B2\rangle$$

$$|2\rangle = |A2\rangle|B3\rangle$$

$$|I\rangle = \frac{1}{\sqrt{2}}(|A3\rangle|B2\rangle + |A2\rangle|B3\rangle)$$

$$|II\rangle = \frac{1}{\sqrt{2}}(|A3\rangle|B2\rangle - |A2\rangle|B3\rangle)$$

Скореллированные пары
 Эйнштейна-Подольского-Розена

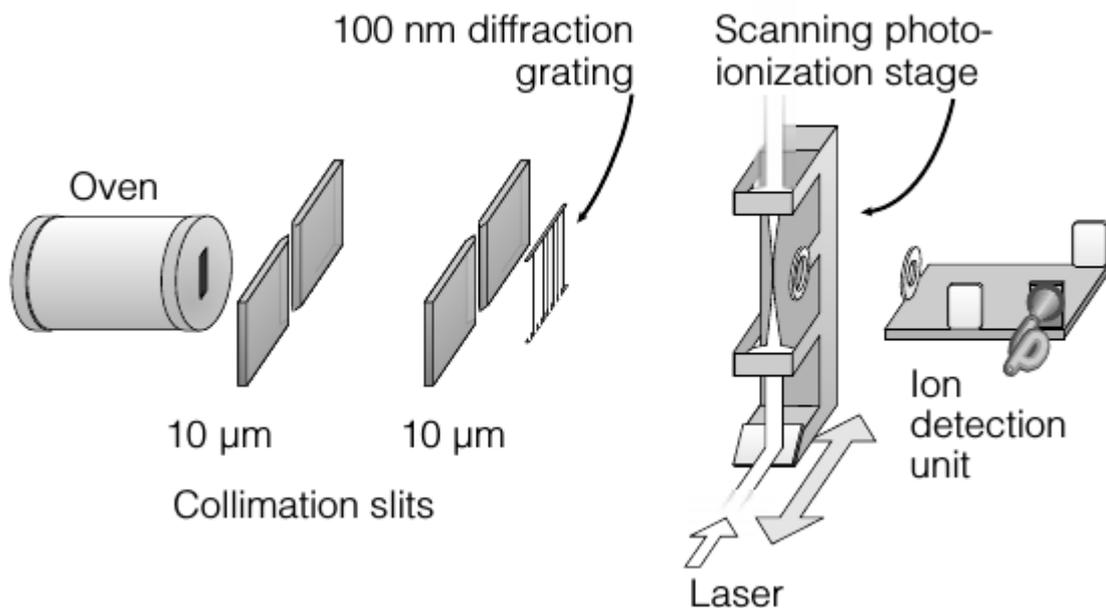
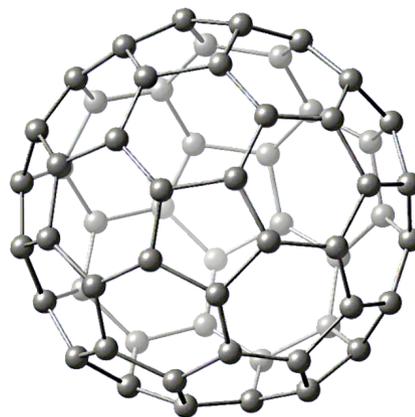
Все зависит от энергетической щели, отделяющей состояния от окружения!

Волновое поведение макромолекул: фуллерен C₆₀

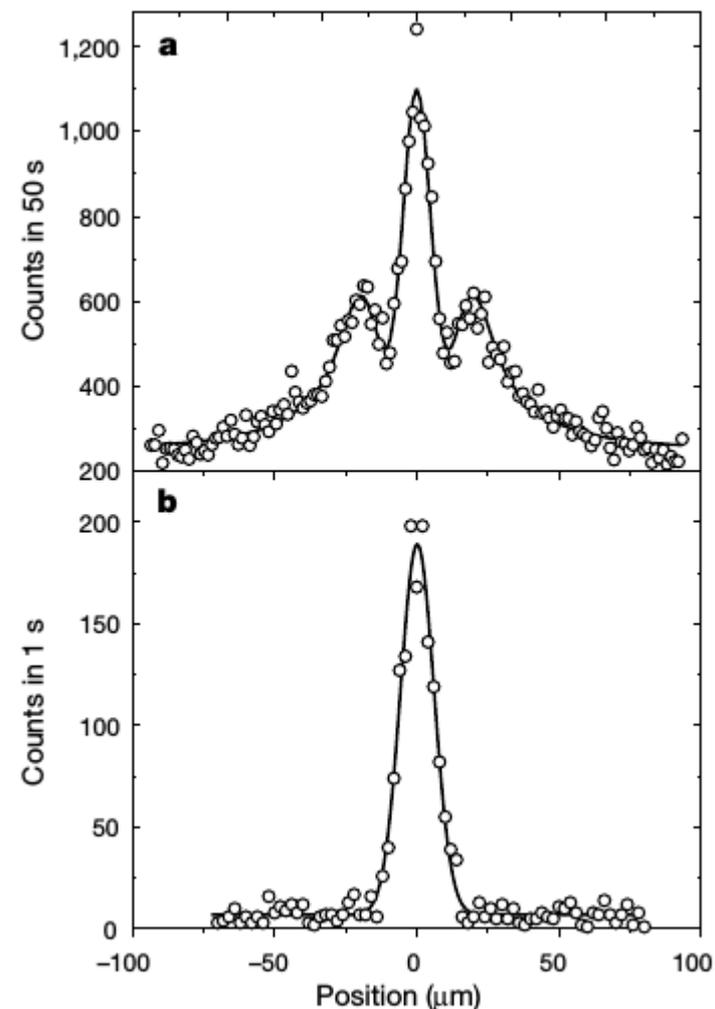
Wave-particle duality of C₆₀ molecules

Markus Arndt, Olaf Nairz, Julian Vos-Andreae, Claudia Keller,
Gerbrand van der Zouw & Anton Zeilinger

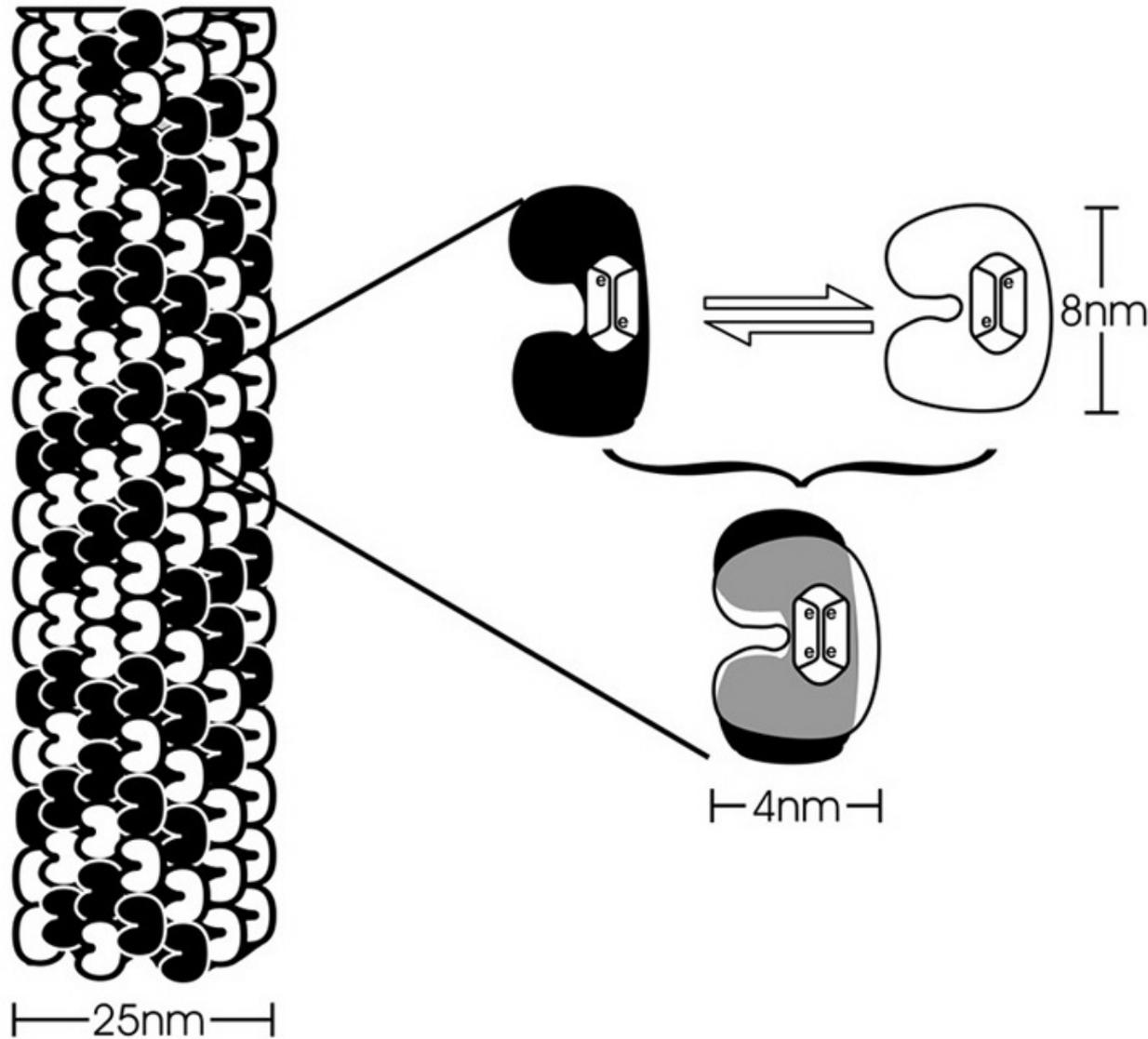
NATURE | VOL 401 | 14 OCTOBER 1999



За время прохождения щелей C₆₀ испускает ~5 фотонов с длиной волны ~10 мкм.
Но размер щелей меньше: 100 нм.



Микротрубочки: Может ли тубулин быть кубитом?



Кристаллическая вода изолирует микротрубочку от внешнего влияния.

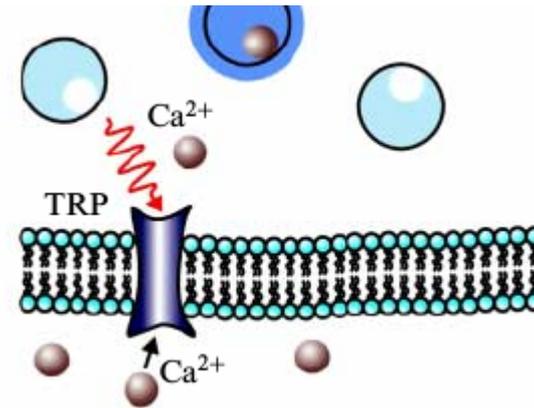
Микротрубочка - квантовый клеточный автомат?

**Может быть да,
Может быть нет.**

Теорема Пенроуза:
В работе мозга **не может не быть** квантовых степеней свободы.

Другие места в клетке, где может иметь место квантово-информационная активность

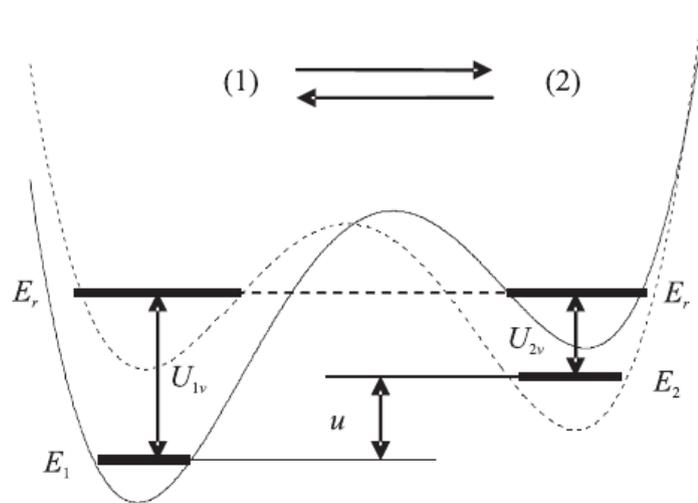
Мембранные ионные каналы.



А.М. Жёлтиков. УФН, Т.188(2018) No 10, С. 1119

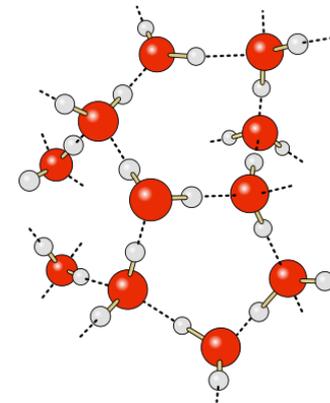
Двухъямный потенциал водородной связи ДНК, РНК, вода (!)

$U(x)$



П. М. Красильников
КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ И
МОДЕЛИРОВАНИЕ Т. 1 (2009) No 3 С. 297

Вода не «простая»,
а «связанная жидкость»



Википедия

Альтернатива нейросетевой парадигме:

Нейросетевая активность не есть носитель сознания, но есть лишь **инструмент** сознания, **интерфейс**, связывающий сознание с окружающей действительностью. «Само» сознание (или его жизненно важные компоненты) живет на некоторых более глубоких (субнейронных, квантовых?) уровнях организации мозга.

Резюме

1. Сколь угодно высокая производительность компьютеров не гарантирует решение задачи создания сильного ИИ.
 - Понимание путей решения задачи не менее важно.
 - Нужна «теория понимания», которой пока нет.
2. Вычислительная производительность мозга, возможно, сильно недооценена. Минимальная мощность компьютеров, необходимых для создания сильного ИИ, (много) выше чем ожидается.
3. Справедливость аналогии «мозг — это компьютер» не доказана. Существенная роль квантовых процессов обработки информации в мозге не исключена.
4. Прогноз создания сильного ИИ к 2045 году обоснован не очень хорошо.

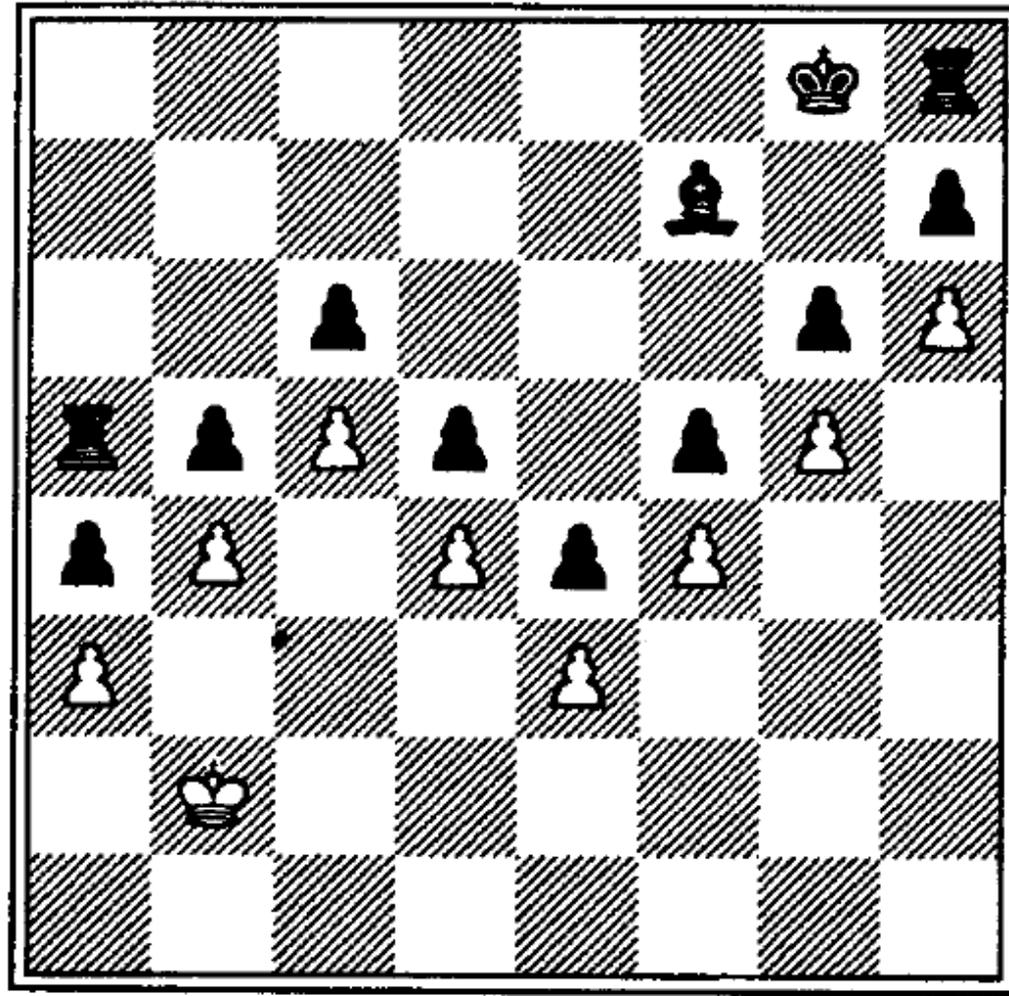


Рис. 1.7. Белые начинают и заканчивают игру вничью — очевидно для человека, а вот «Deep Thought» взял ладью!

Seymore, J., Norwood, D. (1993). A game for life. *New Scientist*, 139, No. 1889, 23–26.