42 Years of Programming

Selected Algorithms and Data Structures

Dipl.Inform. Detlef Wolf

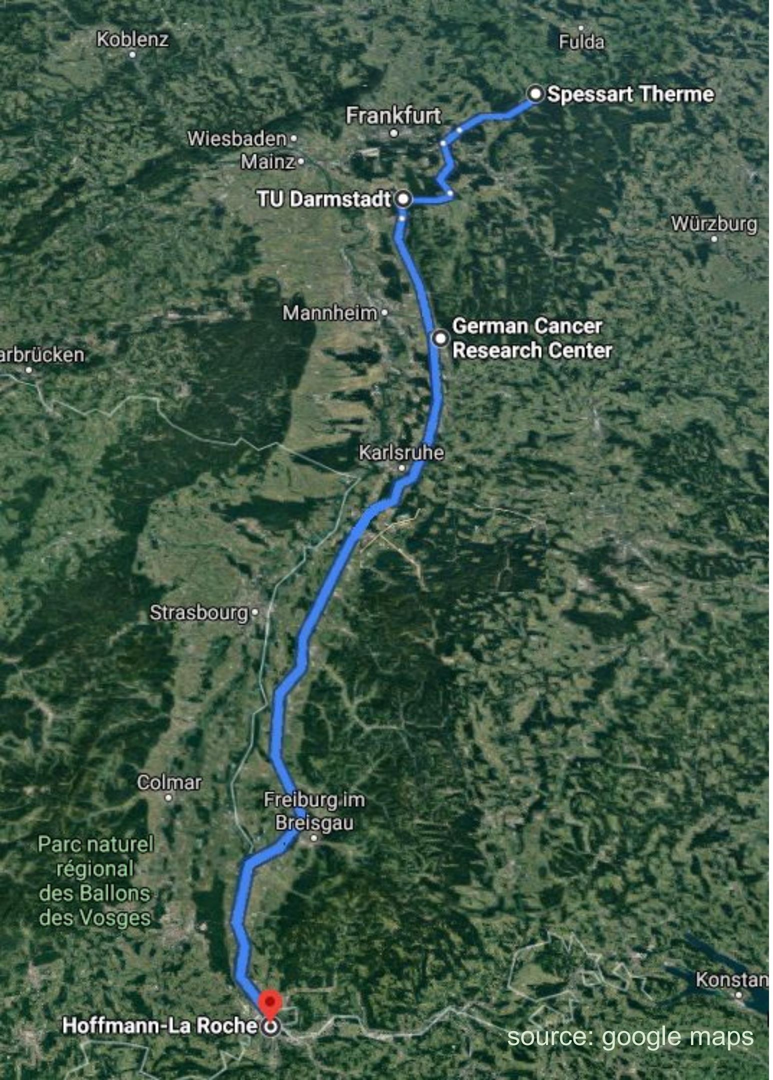
Detlef.Wolf@Roche.com
Roche Innovation Center Basel
F. Hoffmann - La Roche AG

Universität Basel

Applied Mathematics and Informatics in Drug Discovery

https://amidd.ch

4.Dec.2020



source: google maps

Detlef Wolf

1962 * Bad Soden, Hessen, Germany

1980 jugend forscht

1981-1988 Computer Science — Technische Hochschule Darmstadt

1991-1995 dkfz. DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

1996 - Roche

Basel,
New York,
San Francisco

Genetic Algorithm

Fourier Transform

Unitary Matrices

1962 * Bad Soden, Hessen, Germany

1980 jugend forscht

1981-1988 Computer Science — Technische Hochschule Darmstadt

1991-1995 dkfz. DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

1996 - Roche

Basel,
New York,
San Francisco

Abstract Specifications, Floodfill, B*-Trees, Entity-Attribute-Value Schema, Spearman Rank Correlation, ...

Genetic Algorithm - Past

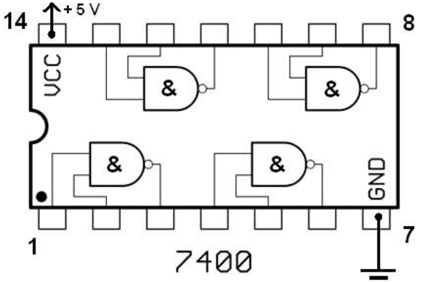
Fourier Transformation - Present

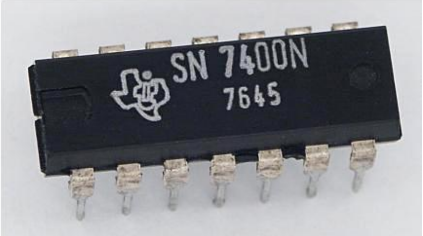
(qu)bit

i^2 = **-1**

Unitary Matrices - Future

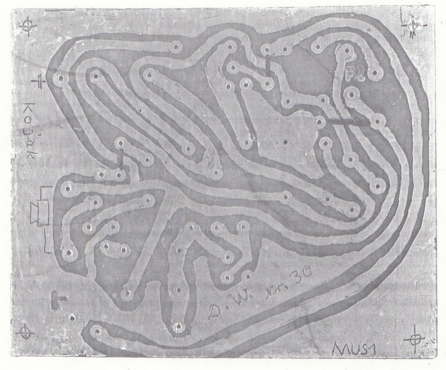
1977

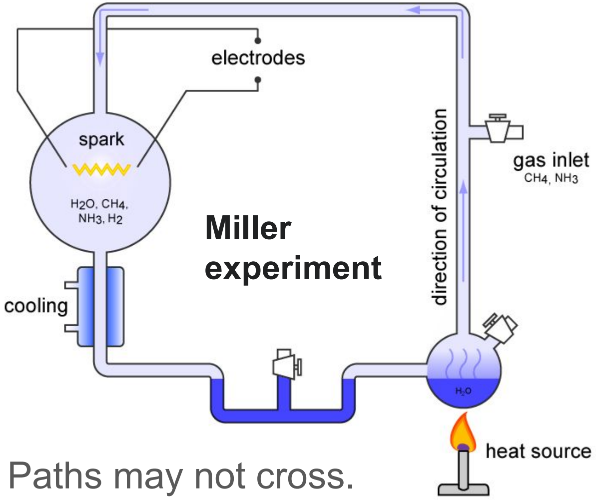




Inputs		Output
A	**B**	**¬(A∧B)**
0	0	1
0	1	1
1	0	1
1	1	0

Charles Darwin





Olivetti P6060

Automatic Printed Circuit Board Design

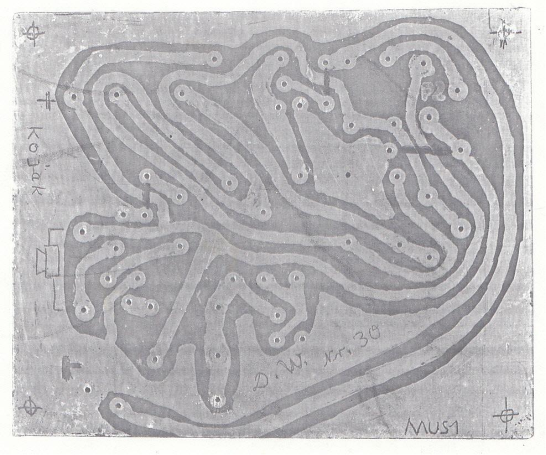
A connection is defined by its two endpoints. Paths may not cross.

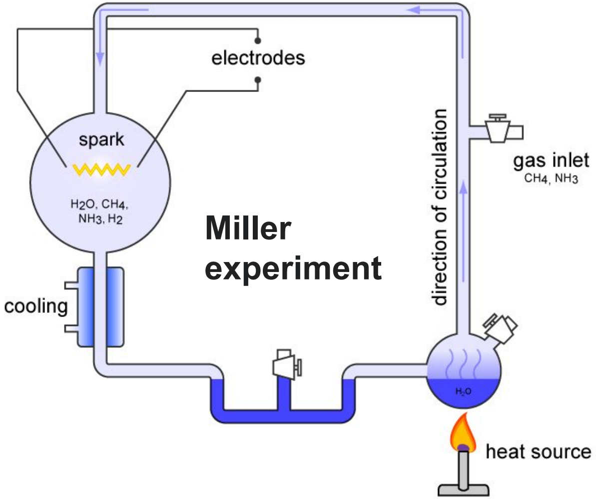
- given N connections to make

1. the algorithm starts with the first connection, route it, then second, …
   - the success depends on the order of making connections
   - there are N! possible orders
2. start with a random order R1, attempt making connections; score: number of connections that could be made plus total length
3. mutate (swap two connections) → R2, try again; if score gets better, continue with R2, else R1
4. Miller experiment (I had understood it as 'lightnings help evolution'): if there is no progress for a while, mutate more, change the order dramatically.

Charles Darwin

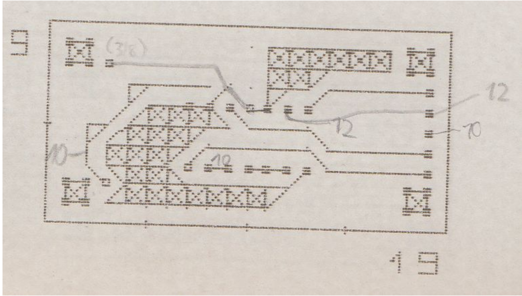
Olivetti P6060





take home message

- **measure it!**
  here: number of connections & total length
- **build expectations!**
  have an idea of what you expect as outcome … and if there is something unexpected explore it!
  here: evolution into dead end



Now comes a big jump in time ...

Parkinson's Disease (1)

- **Key Descriptions:**
  - in 1817, James Parkinson (British physician): An Essay on the Shaking Palsy
  - in 1912, Friedrich Lewy described abnormal aggregates of protein in PD patients
  - Dopaminergic cell loss within the substantia nigra pars compacta of the brain

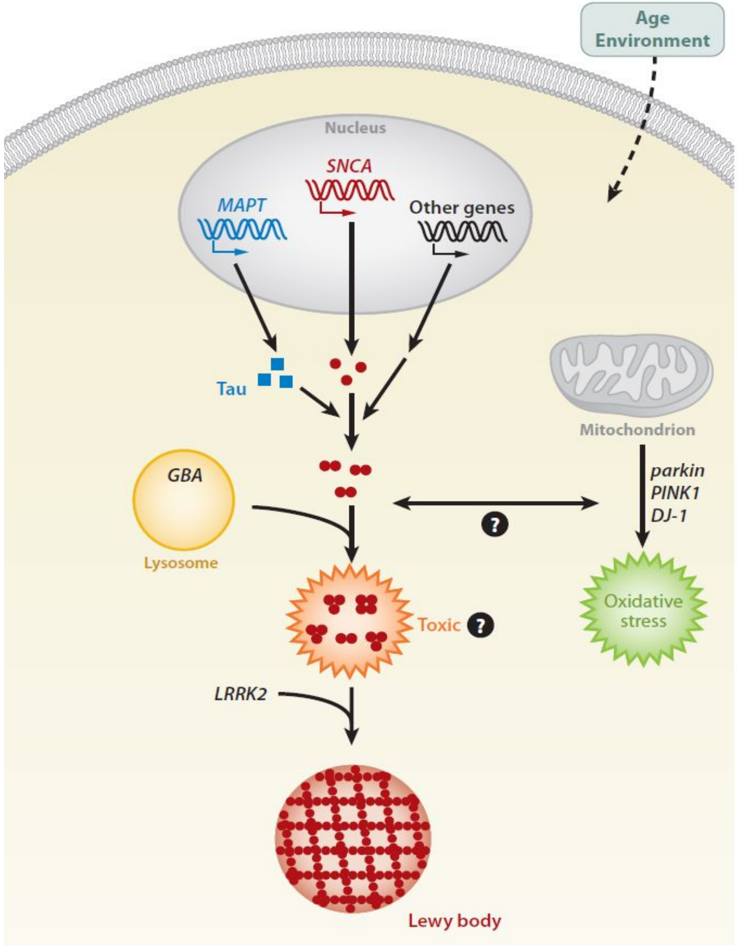
- **Symptoms:**
  - Trembling of hands, arms, legs, jaw and face
  - Slowness of movement (walking)
  - Stiffness of the arms, legs and trunk

- **Progression**:
  onset around age 60; chronic and slowly progressive disorder with a mean duration of 15 years from disease recognition until death

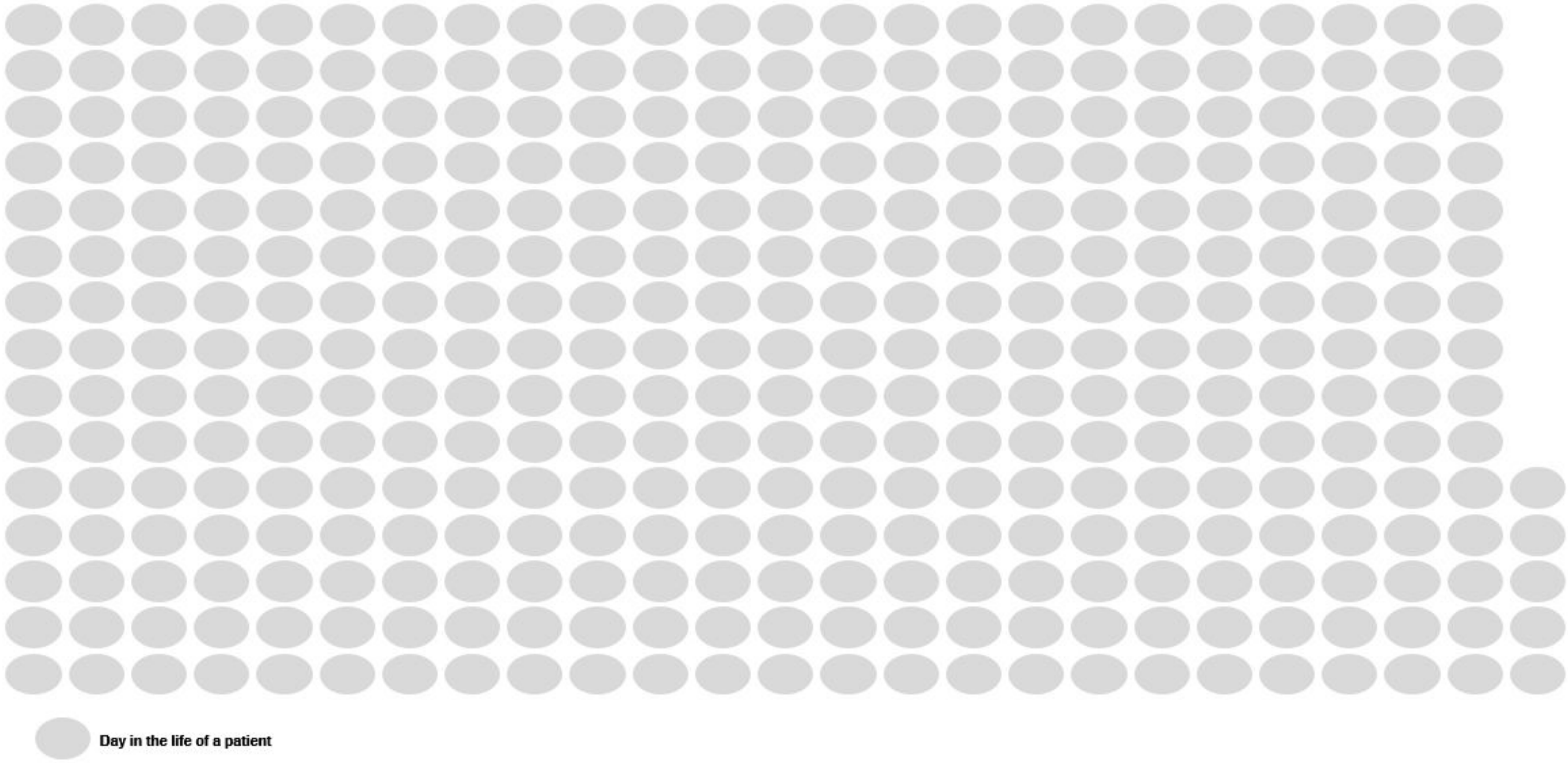
Parkinson’s Disease (2)

- **Diagnosis:**
  - no established lab test only
  - extensive neurological tests
  - PET scan starting to be used
- **Therapy:**
  - symptomatic : now
  - disease-modifying – future?
- **Hypothesis**:

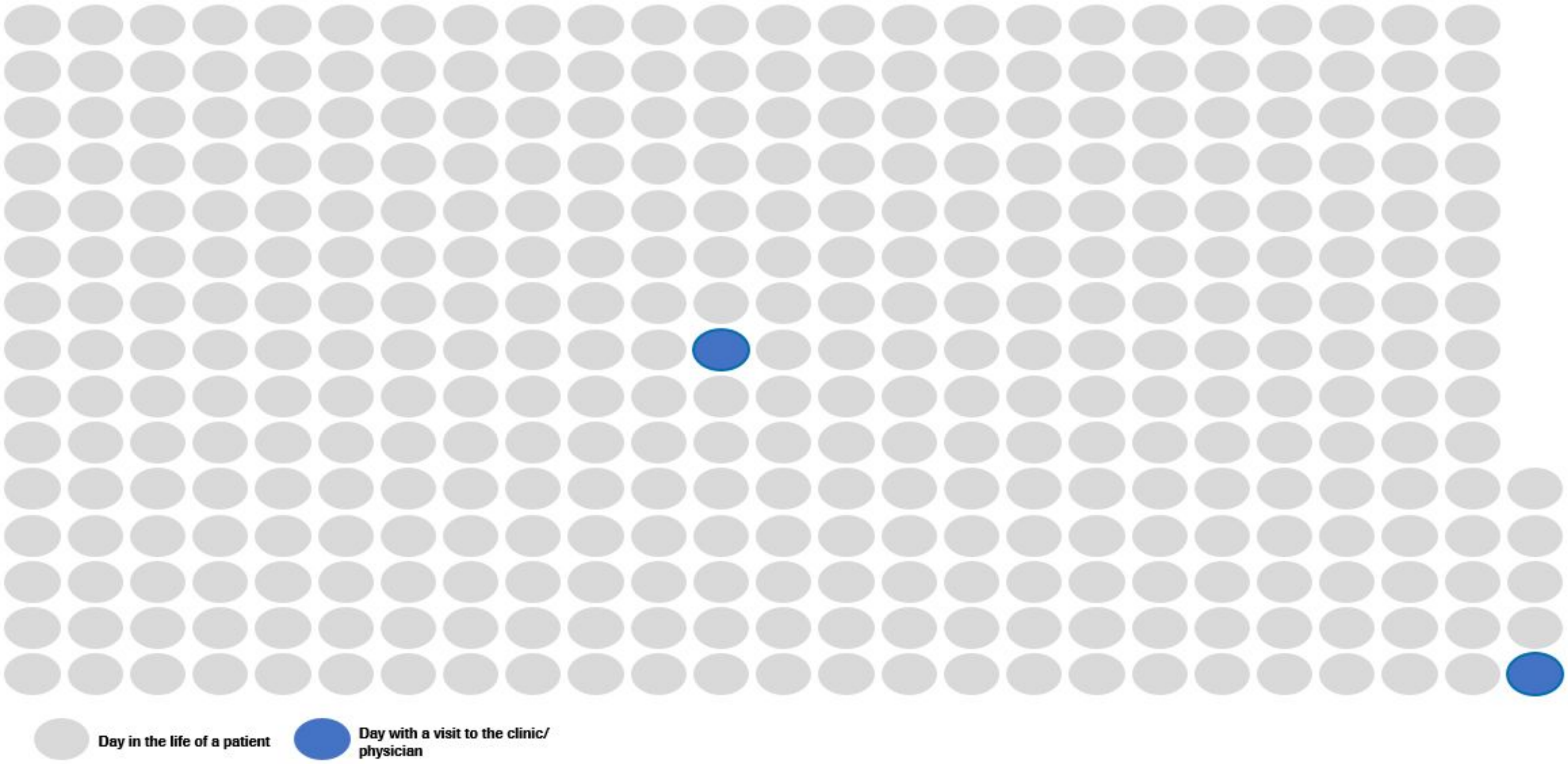
  Aggregation of α-synuclein (product of gene SNCA) is proposed as the central pathway leading to neurotoxicity in PD.
  Other hypotheses include: PINK1, PARK2, PARK7, LRRK2



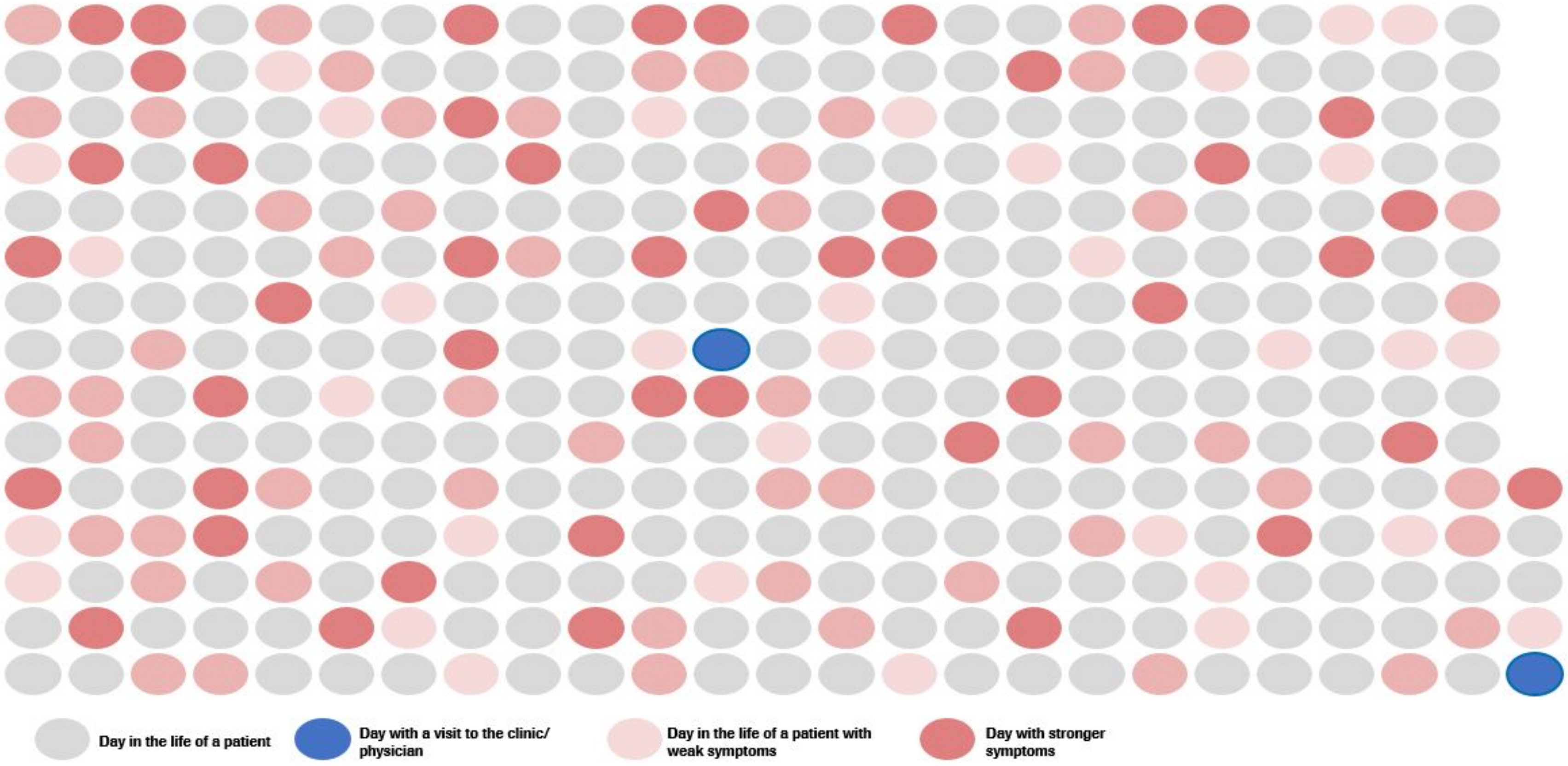
Remote patient monitoring can transform clinical research



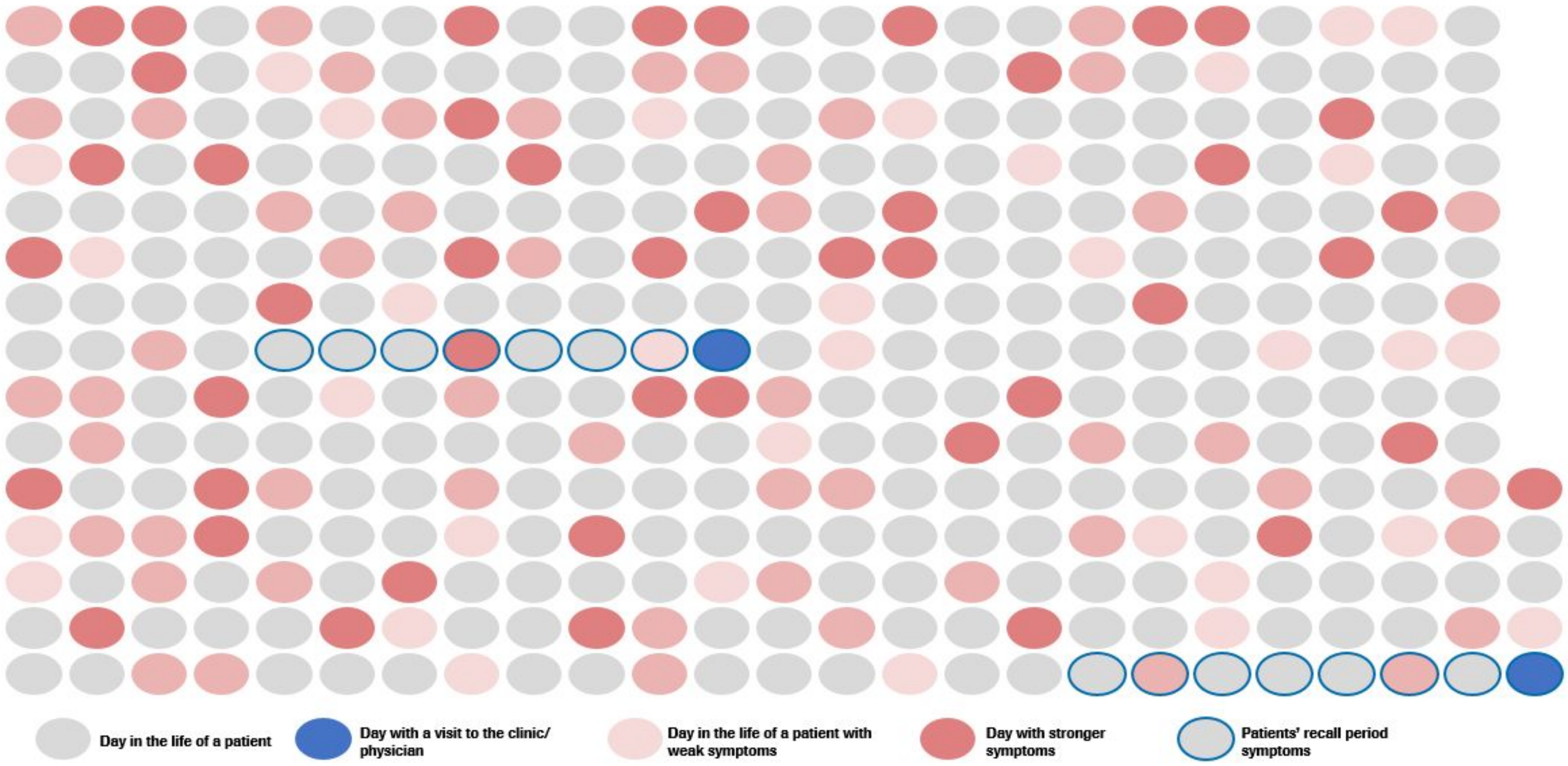
Remote patient monitoring can transform clinical research



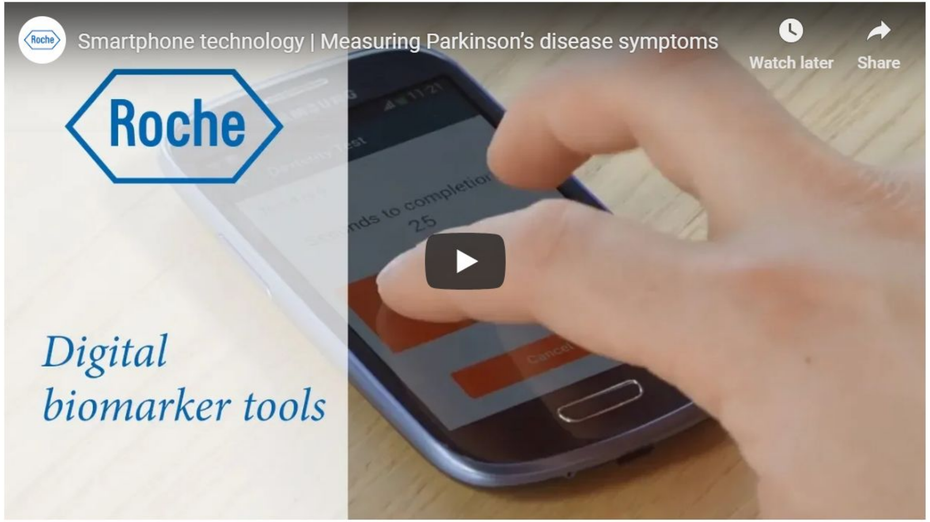
Remote patient monitoring can transform clinical research

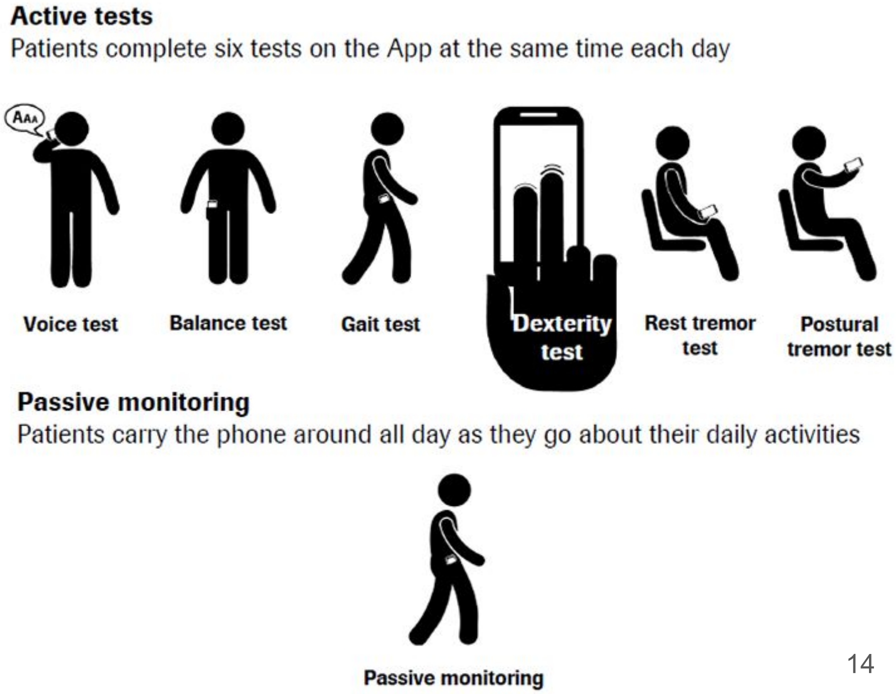


Remote patient monitoring can transform clinical research



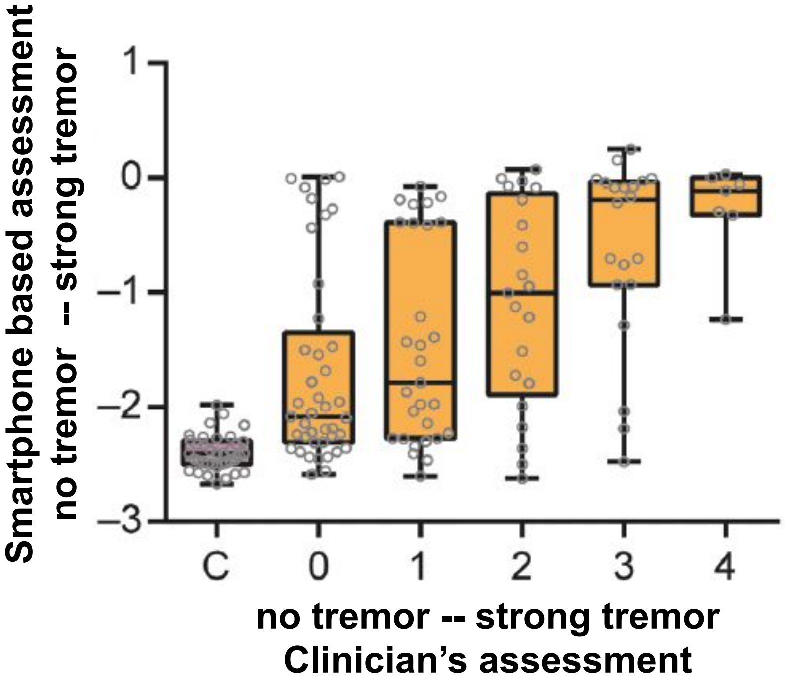
Roche technology measures Parkinson's disease fluctuations





https://youtu.be/qAu9SyfQsQY

It is possible to measure Parkinson's Disease status with smartphones

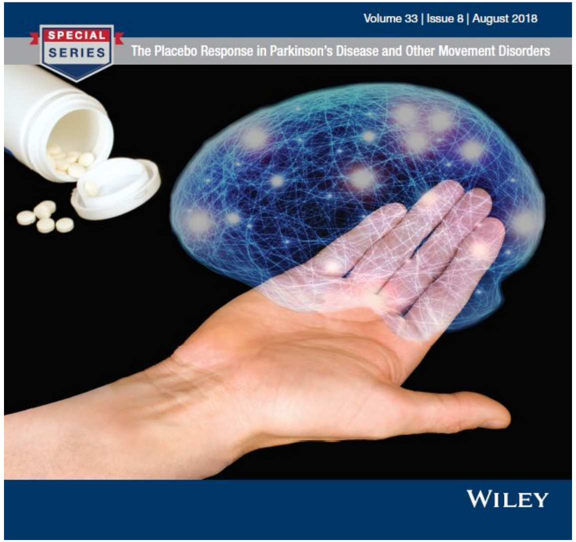


Official Journal of the International Parkinson and Movement Disorder Society

Movement Disorders

Roche

2018



RESEARCH ARTICLE

Evaluation of Smartphone-Based Testing to Generate Exploratory Outcome Measures in a Phase 1 Parkinson's Disease Clinical Trial

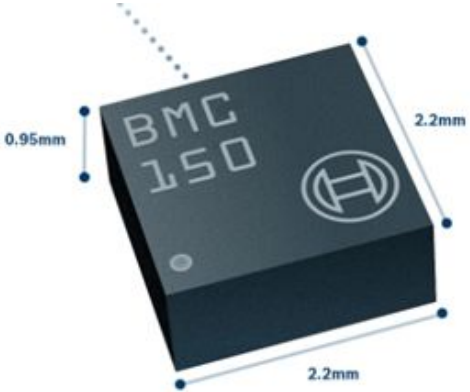
PHASE 1 Study

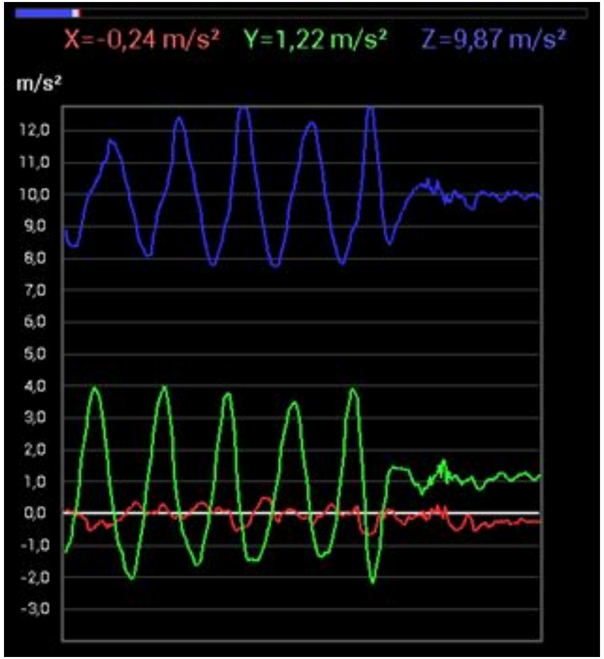
Potential to shorten clinical trials

Example Device and Sensor

If you would like to try out things yourself :

https://phyphox.org/





A digital biomarker learning experience

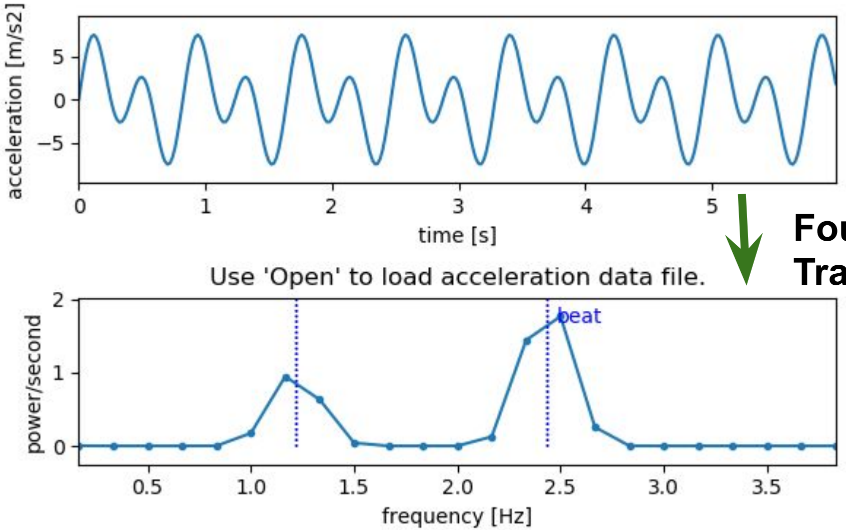
For a conference, we wanted to make a digital biomarker learning experience with a live element.
To add a fun element, instead of measuring tremor, we measured dancing -- after all it (should be) also a periodic movement, thus analyzable be a Fourier transform.

Digital biomarkers are **defined** as objective, quantifiable physiological and behavioral data that are collected and measured by means of **digital** devices such as portables, wearables, implantables or digestibles. The data collected is typically used to explain, influence and/or predict health-related outcomes.

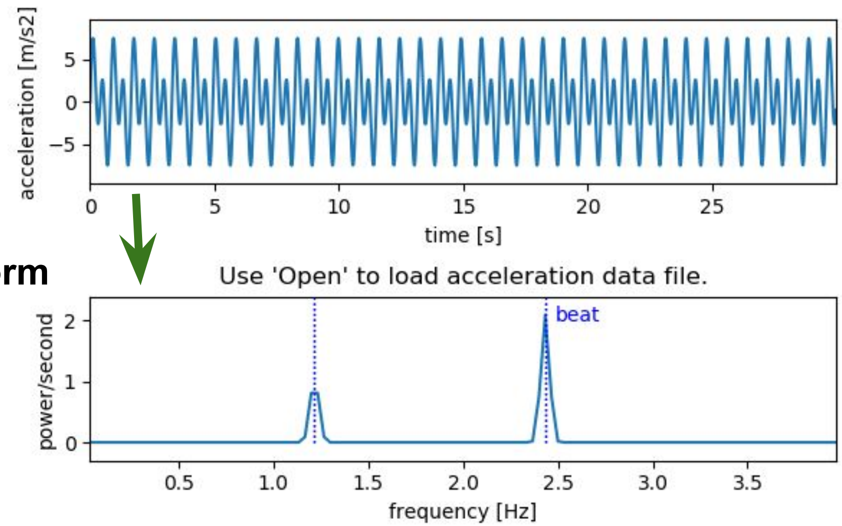
we used this song:
https://www.youtube.com/watch?v=KXO_SIg5bV0

Simulate Data, Check Correctness of Software & Understanding

```python
def generateDummy():
    bfs = 146.0 / 60.0   # beat frequency [Hz] (146 beats per min.)
    bfs2 = bfs / 2.0     # [Hz]
    sampling_frequency = 1 / 0.02 # Hz
    duration = 30.0 # seconds
    timepoints = np.arange(0.0, duration, 1.0/sampling_frequency)
    acc = (np.sin(2.0 * np.pi * bfs * timepoints) +
           np.sin(2.0 * np.pi * bfs2 * timepoints) /
           math.sqrt(2)) * 9.81 / 2.0
    return sampling_frequency, timepoints, acc
```



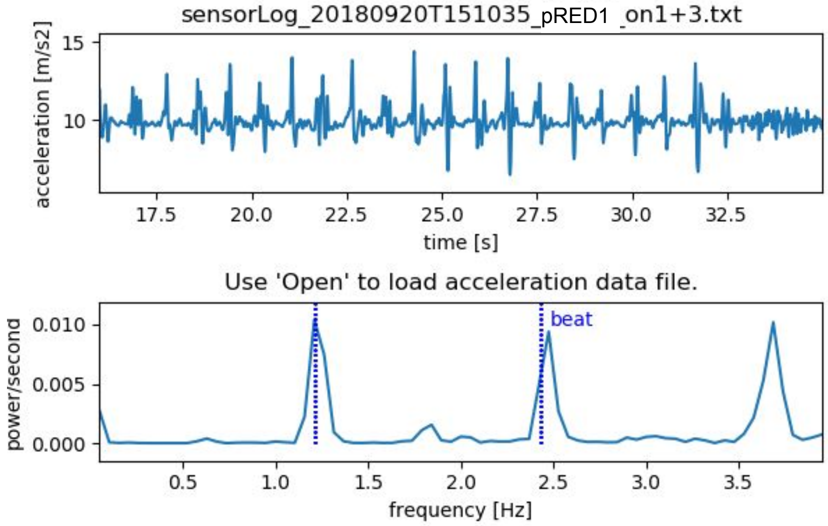
Fourier Transform

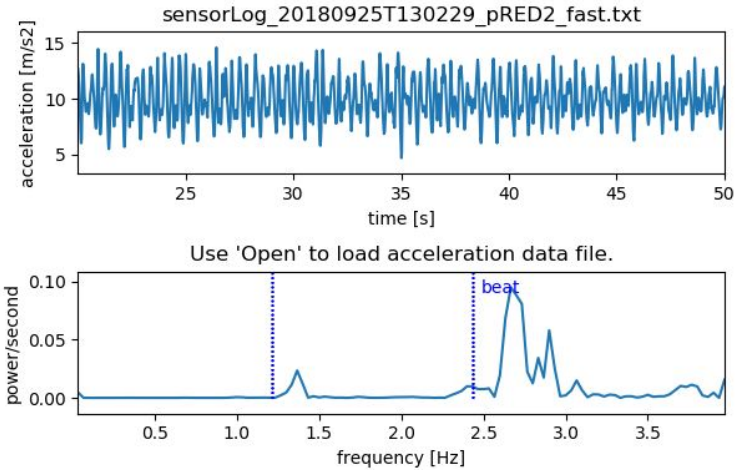


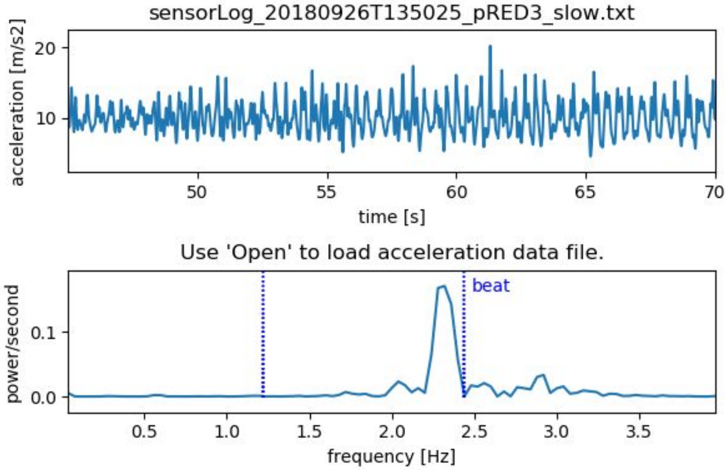
→ 30 seconds of sampling can give nice peaks (left), at least 6 seconds it should be (right)

Measure some persons dancing

Yeah, it works!
one person dancing at the right speed,
another too fast, another too slow … looks reasonable

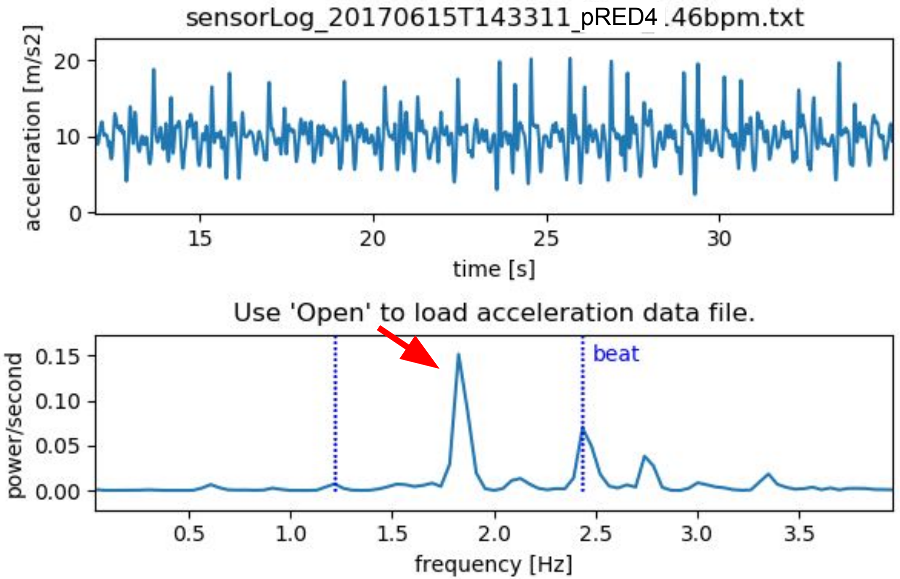


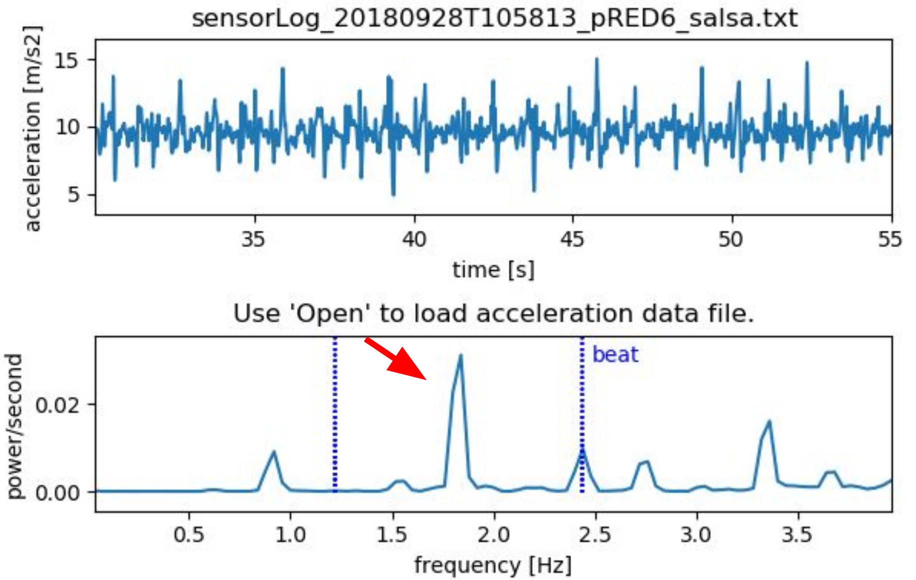


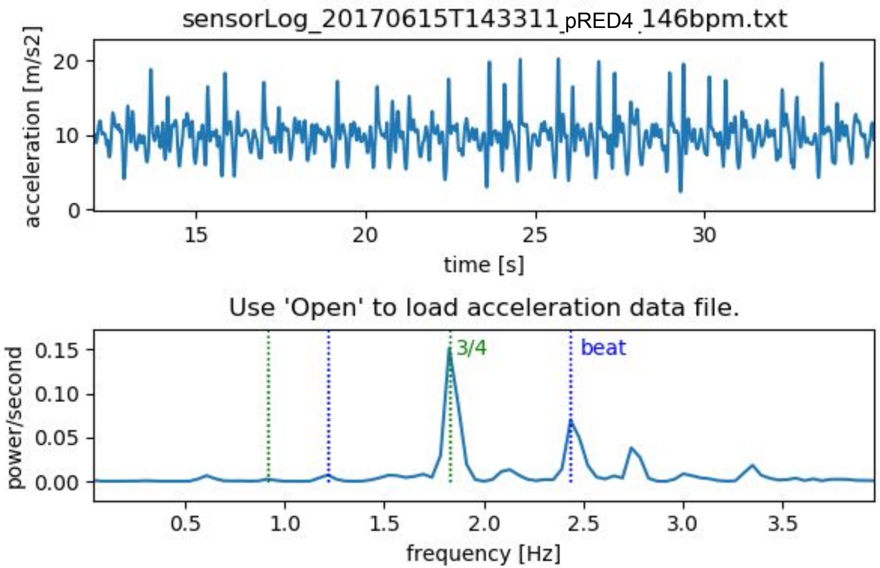


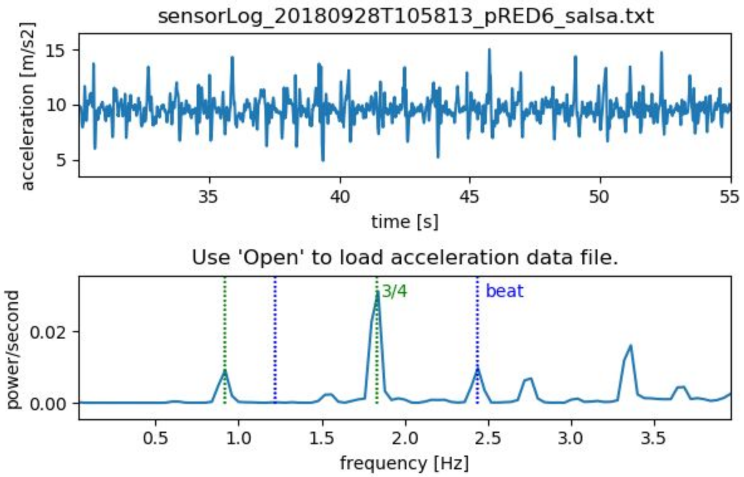
Measure more persons

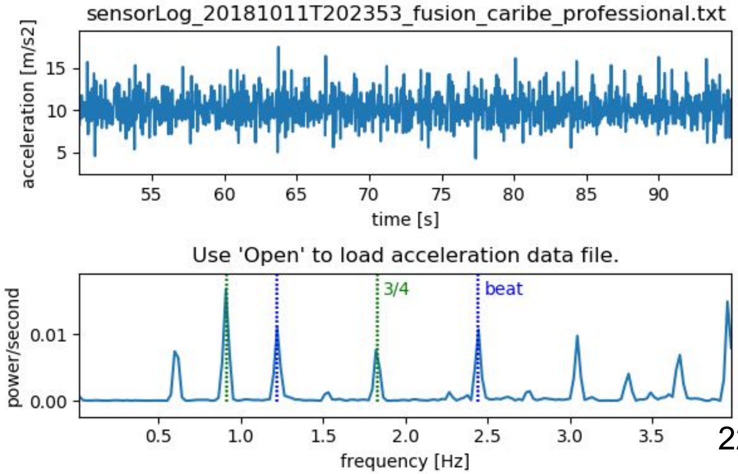
But what are these strange peaks in the middle? (samples from different persons)









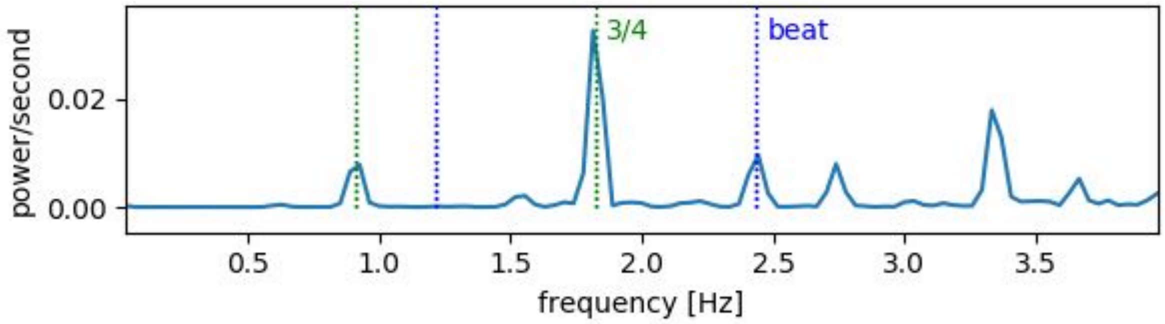


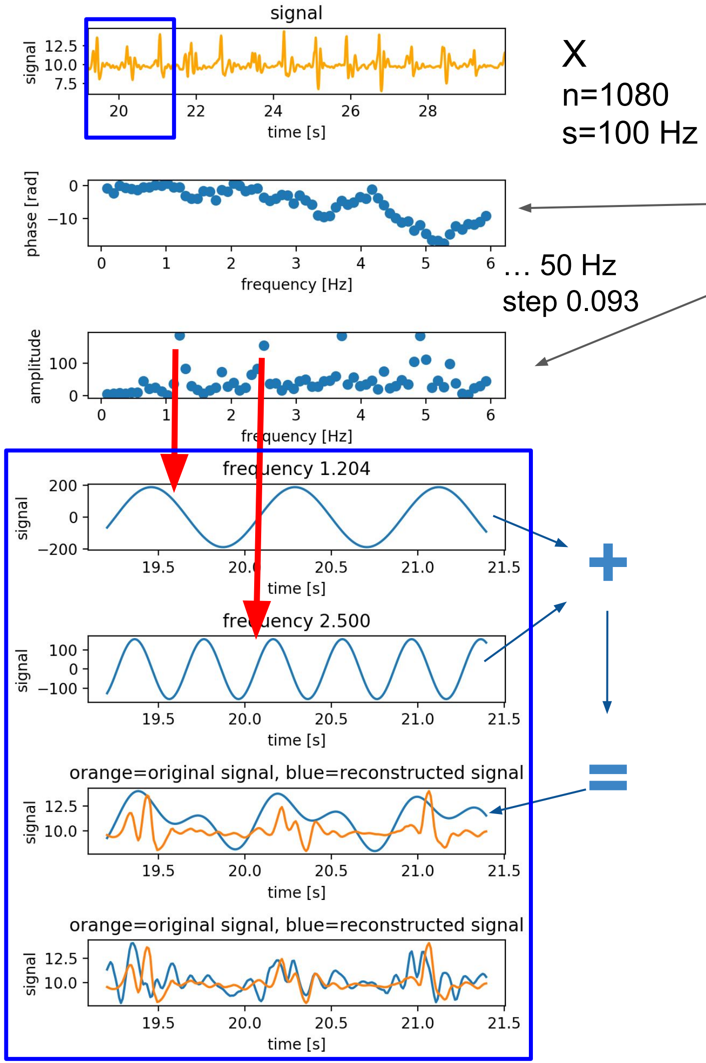
Measuring their position shows they are at 3/4 of the beat frequency.

We assume this is due to the nature of this dance (Salsa): there are **3 steps on 4 beats**, producing a **superposition** of two rhythms.
Lower right: professional Salsa dancer

take home message (again)

- **measure it!**
  here: location of peaks
- **build expectations!**
  have an idea of what you expect as outcome …
  and if there is something unexpected explore it!
  here: only peak at beat or its harmonics





Discrete Fourier Transform

transform from the time domain into the frequency + phase domain

Y = numpy.fft.rfft(X) **[adapted from Python manual]**

This function computes the one-dimensional *n*-point discrete Fourier Transform (DFT) of a real-valued vector X by means of an efficient algorithm called the Fast Fourier Transform (FFT).

s = sampling frequency, n = length(X)

Y → n/2 complex numbers,
one per frequency f: 0 … s/2 in steps of s/n
Each complex number *contains* amplitude
and phase information

amp = sqrt(imag(f)^2 + real(f)^2) → complex absolute (modulus)

phase = atan2(imag(f),real(f))*180/pi

Why do today's discrete fourier transform implementations return complex numbers?

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos(2\pi n x) + b_n \sin(2\pi n x))$$

Original Fourier Series: only real numbers

$$e^{i\varphi} = \cos\varphi + i\sin\varphi$$

Euler's formula

$$\sqrt{-1} = i \qquad i^2 = -1$$

exponential representation allows more efficient computation

$$c = a + ib$$

a, b are real numbers

$$|c| = +\sqrt{a^2 + b^2}$$

modulus of complex number

Quantum computers promise to offer computing power far beyond anything we can imagine. They have enormous potential, especially in the pharmaceutical industry. That is why we have a task force that is preparing for "the big wave".

https://www.roche.com/quantum-computing.htm

Dec 2018

Goals in Quantum Computing (QC) at Roche

Chemistry

Optimization

Machine Learning

Will QC disrupt X?

When will QC disrupt X?

What are the requirements for QCs to disrupt X?

Accurate binding affinity estimation would strongly influence efficiency in **chemical Lead Identification/Optimization**.
Rapid identification of minima would broadly influence Research & Early Development.
If Machine Learning is improved by QC, better predictions would impact areas from **biological target assessment to in-vivo trials** and beyond.

1 Qubit

state

"how much zero"

"how much one"

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

allows Superposition while not measured /

binary when measured

$$|\alpha|^2 + |\beta|^2 = 1$$

$$\alpha = a + ib$$

$$|\alpha| = +\sqrt{a^2 + b^2}$$

probability of measuring "one":

$$p(|1\rangle) = \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2} = |\beta|^2$$

alternate representation:

$$|\Psi\rangle = \cos(\theta)|0\rangle + e^{i\phi}\sin(\theta)\,|1\rangle$$

There is no way of measuring **α** or **β** directly. When measuring the qubit shows either 0 or 1, nothing in between.

1 Qubit example

$$|\Psi_1\rangle = \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} = \overset{\alpha}{\frac{1}{\sqrt{2}}}\begin{bmatrix}1\\0\end{bmatrix} + \overset{\beta}{\frac{1}{\sqrt{2}}}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

one qubit in 50:50 superposition

Dirac's notation

Notational Forrest

$$p(|1\rangle) = \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2} = 0.5$$

apply a 1-qubit operator

Hadamard

q(t)

q(t+1)

$$\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\1/\sqrt{2} & -1/\sqrt{2}\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$

from pure zero to 50:50 superposition

a unitary matrix

From 1 qubit to 2 qubits
Preparation

There a several ways to multiply vectors ...

1 number

Matrix multiplication $$\begin{bmatrix} a & c \end{bmatrix} * \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} ae + cf \end{bmatrix}$$

tensor product $$\begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \begin{bmatrix} e \\ f \end{bmatrix} \\ b \begin{bmatrix} e \\ f \end{bmatrix} \end{bmatrix} = \begin{bmatrix} ae \\ af \\ be \\ bf \end{bmatrix}$$

4 numbers

2 Qubits

combining two qubits

$$|\Psi_2\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix} \otimes \begin{bmatrix}\gamma\\ \delta\end{bmatrix} = \begin{array}{c} \begin{bmatrix}\alpha\gamma\\ a\delta\\ \beta\gamma\\ \beta\delta\end{bmatrix} \end{array} \begin{array}{c} |00\rangle\\ |01\rangle\\ |10\rangle\\ |11\rangle \end{array}$$

apply a 2-qubit operator

$$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\alpha\gamma\\ a\delta\\ \beta\gamma\\ \beta\delta\end{bmatrix} = \begin{bmatrix}\alpha\gamma\\ a\delta\\ \beta\delta\\ \beta\gamma\end{bmatrix}$$

swap (C-NOT)

another unitary matrix

combining 1-qubit quantum operators

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} * \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae + cf & ag + ch \\ be + df & bg + dh \end{bmatrix}$$

m=number of qubits

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \otimes \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} a\begin{bmatrix} e & g \\ f & h \end{bmatrix} & c\begin{bmatrix} e & g \\ f & h \end{bmatrix} \\ b\begin{bmatrix} e & g \\ f & h \end{bmatrix} & d\begin{bmatrix} e & g \\ f & h \end{bmatrix} \end{bmatrix} = \begin{array}{cc} \mathbf{2^m} & \\ \begin{bmatrix} ae & ag & ce & cg \\ af & ah & cf & ch \\ be & bg & de & dg \\ bf & bh & df & dh \end{bmatrix} & \mathbf{2^m} \end{array}$$

Transition Matrices for two 1-qubit operations

tensor product

Transition Matrix for one 2-qubit operation

2-qubit quantum operator: non separable → entanglement

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} * \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae + cf & ag + ch \\ be + df & bg + dh \end{bmatrix}$$

m=number of qubits

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \otimes \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} a\begin{bmatrix} e & g \\ f & h \end{bmatrix} & c\begin{bmatrix} e & g \\ f & h \end{bmatrix} \\ b\begin{bmatrix} e & g \\ f & h \end{bmatrix} & d\begin{bmatrix} e & g \\ f & h \end{bmatrix} \end{bmatrix} = \begin{array}{c} 2^m \\ \begin{bmatrix} ae & ag & ce & cg \\ af & ah & cf & ch \\ be & bg & de & dg \\ bf & bh & df & dh \end{bmatrix} \end{array} 2^m$$

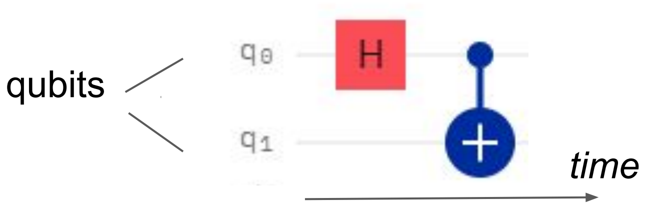
Transition Matrices for two 1-qubit operations

tensor product

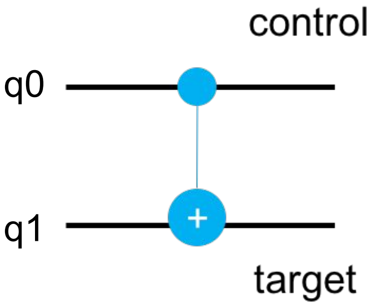
Transition Matrix for one 2-qubit operation

C-NOT is not a tensor product!

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



A quantum computing program



Conditional NOT

Hadamard q0(t) q0(t+1)

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

q0 q1

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|10\rangle$$

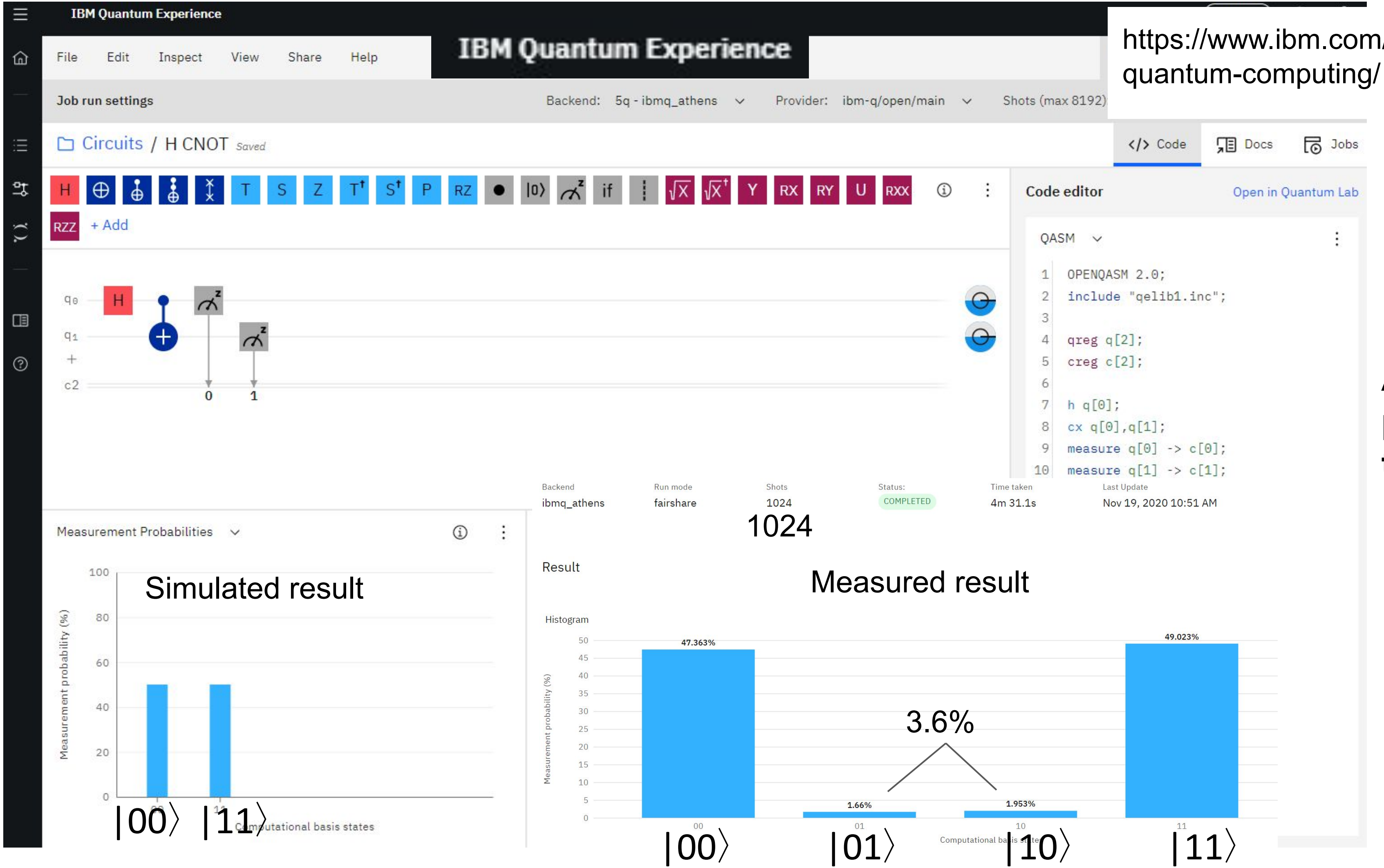
Combine with second qubit

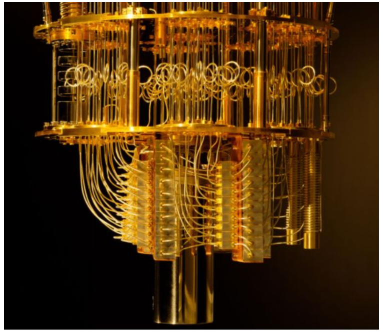
C-NOT

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{array}{cc} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{bmatrix} & \begin{array}{c} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{array} \end{array}$$

https://www.ibm.com/quantum-computing/

A practical test





take home message (once again)

- measure it!

here: get access to a quantum computer

- build expectations!

have an idea of what you expect as outcome …

and if there is something unexpected explore it!

here: future will tell ...

Acknowledgements

Clemens Broger

Martin Strahm & the Roche Quantum Computing Team

Florian Lipsmeier & the Digital Biomarker Data Analysis Team

Xing Chen

David Jitao Zhang

Thank you for your attention.
Questions welcome.