ICEPP
The University of Tokyo

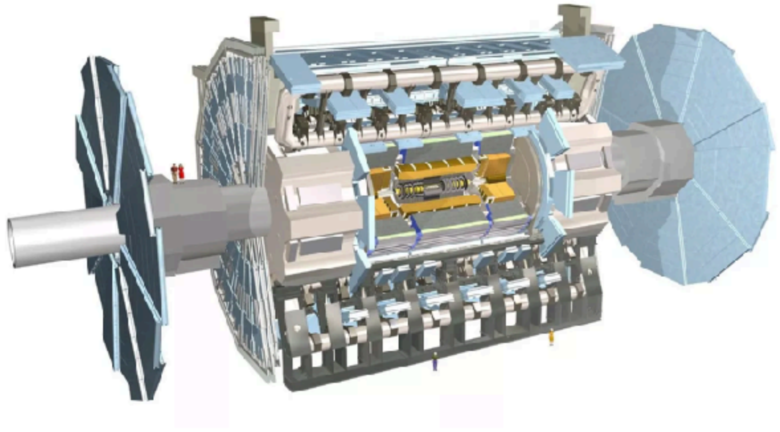
Advanced computing and software for HEP

Yutaro Iiyama
ICEPP, The University of Tokyo

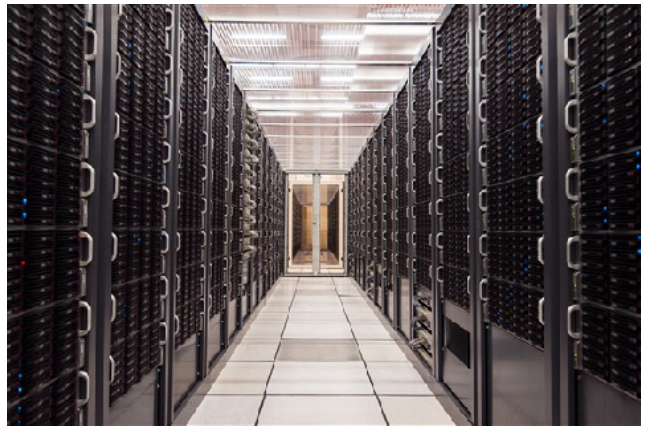
Windows on the Universe
30th Anniversary of the Rencontres de Vietnam

Computing is integral to HEP

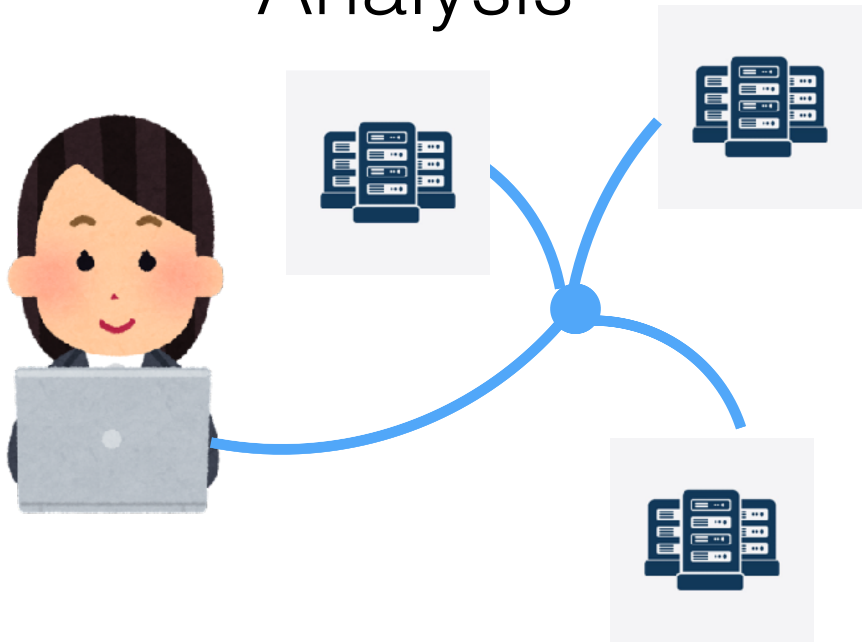
Data acquisition



Central processing /
Simulation

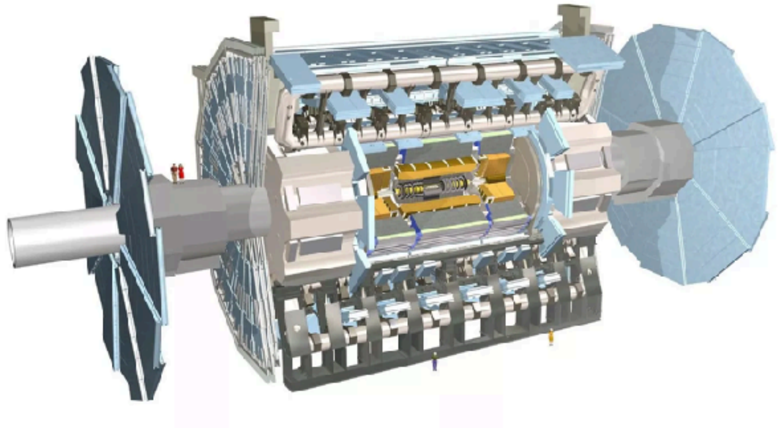


Analysis

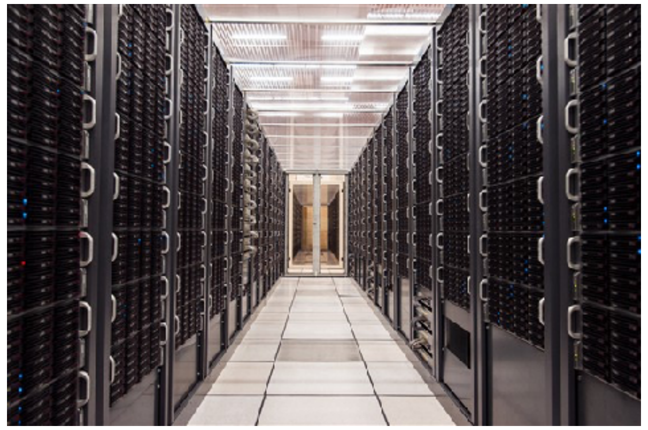


Computing is integral to HEP

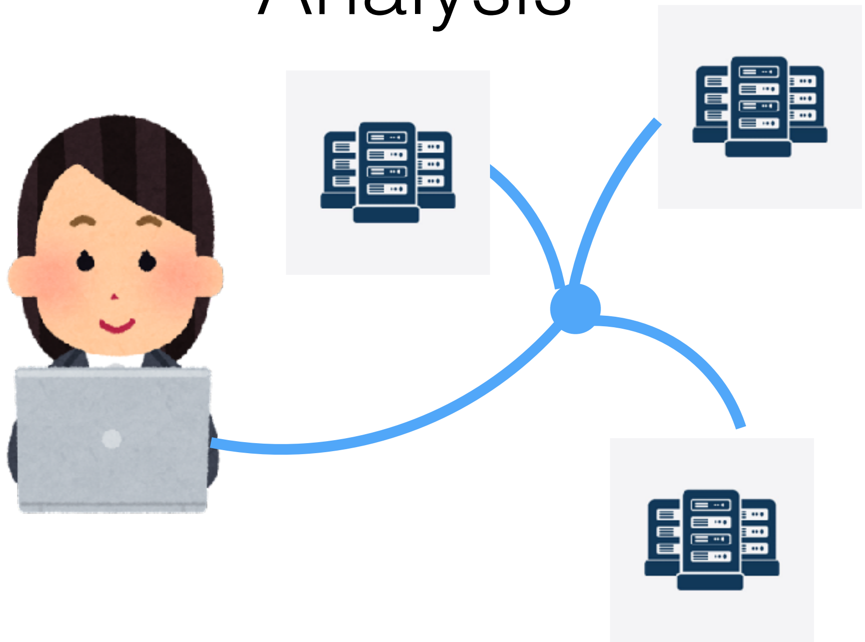
Data acquisition

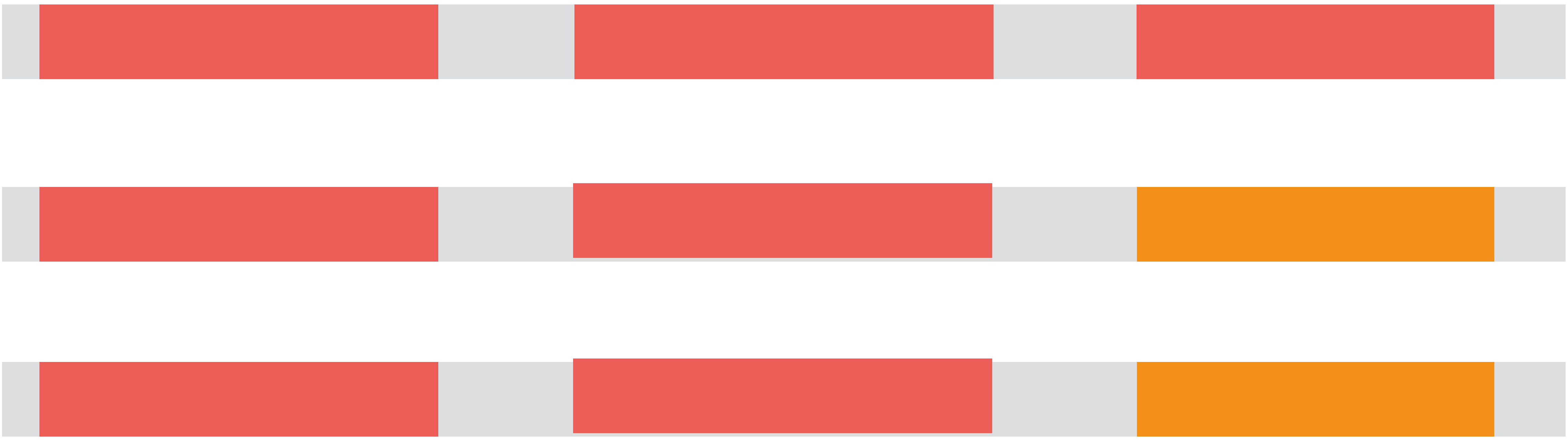


Central processing / Simulation



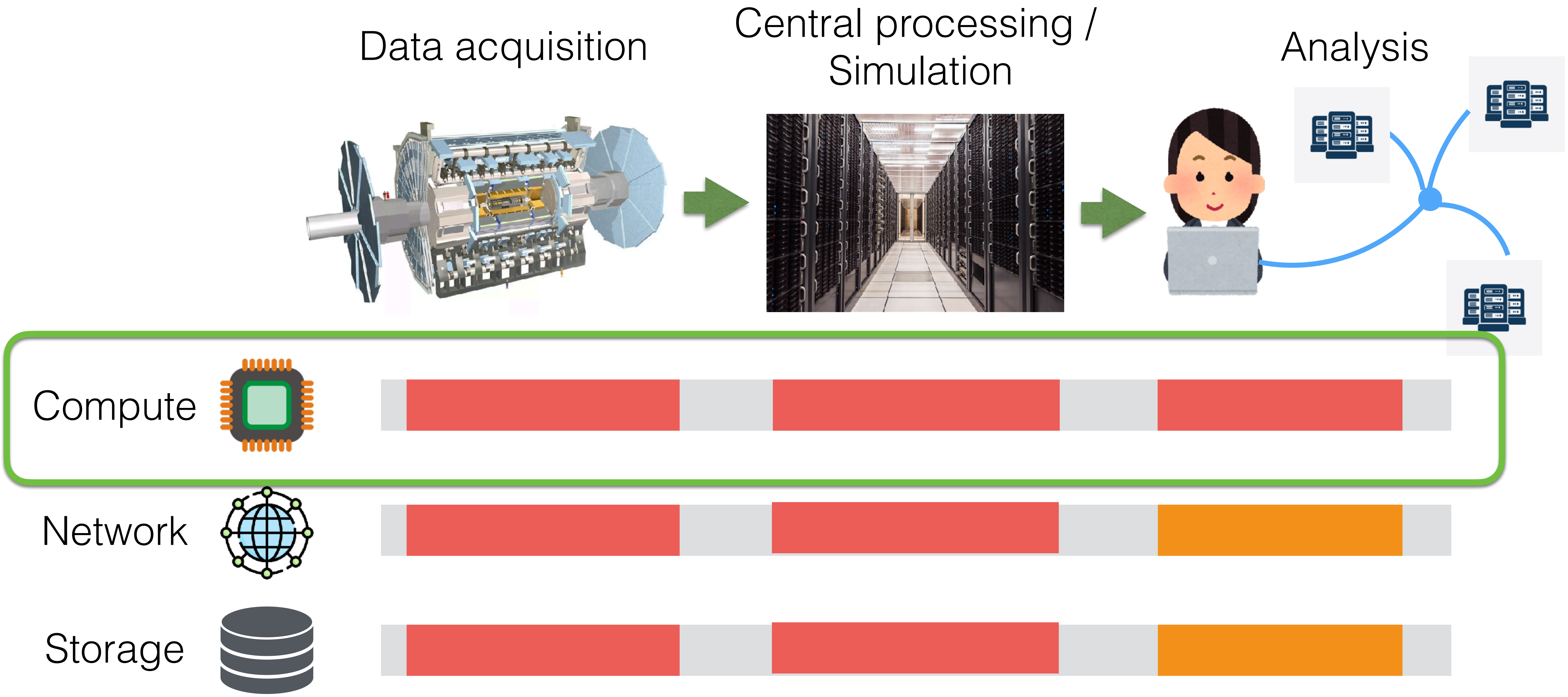
Analysis





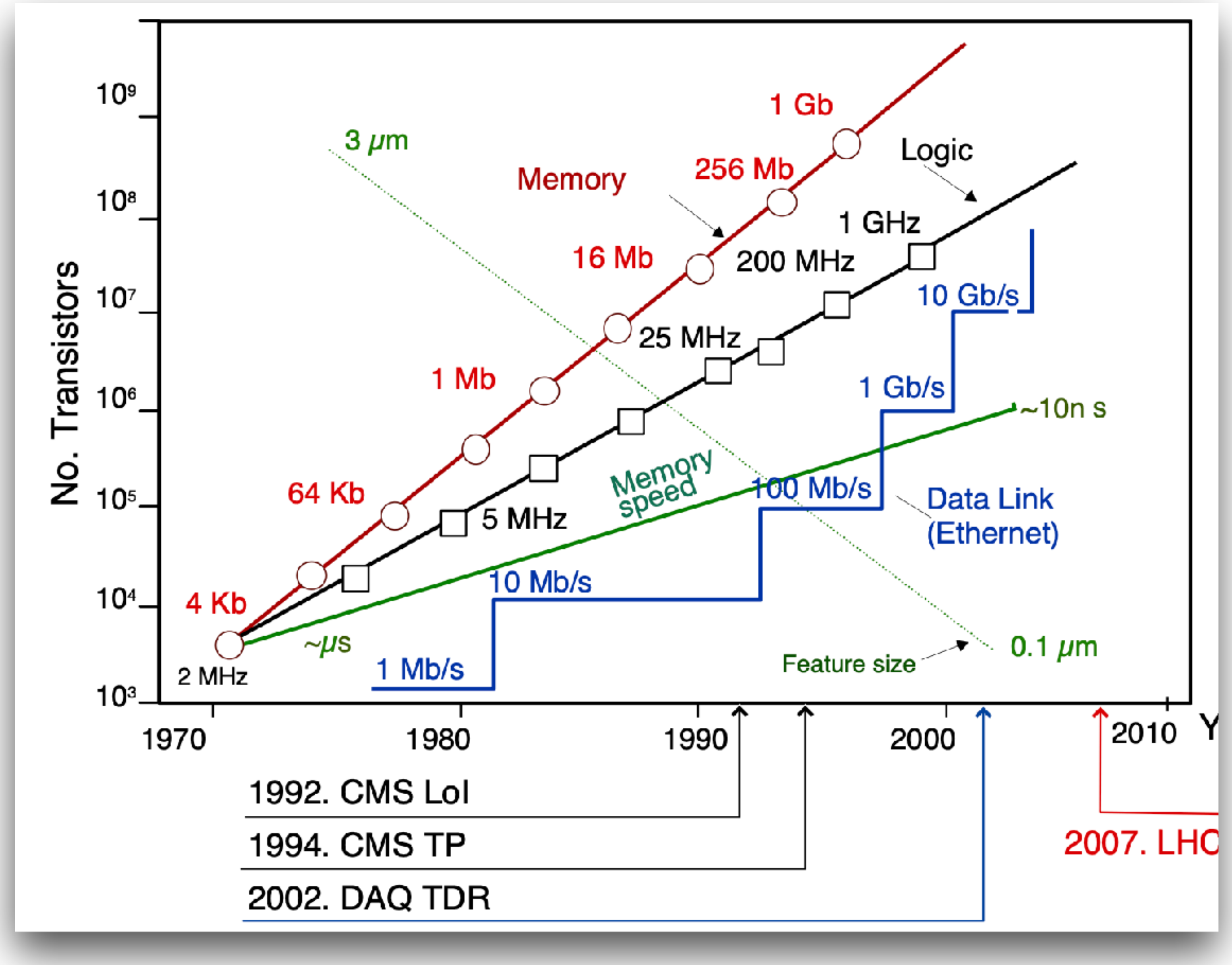
	Data acquisition	Central processing / Simulation	Analysis
Compute	🟥	🟥	🟥
Network	🟥	🟥	🟧
Storage	🟥	🟥	🟧

Computing is integral to HEP



	Data acquisition	Central processing / Simulation	Analysis
Compute	🟥	🟥	🟥
Network	🟥	🟥	🟧
Storage	🟥	🟥	🟧

Investment in advanced computing pays off



CMS Data Acquisition and
High-Level Trigger TDR
(2002)

Example: CMS two-tiered trigger

Hardware (L1T) → CPU farm (HLT)
when

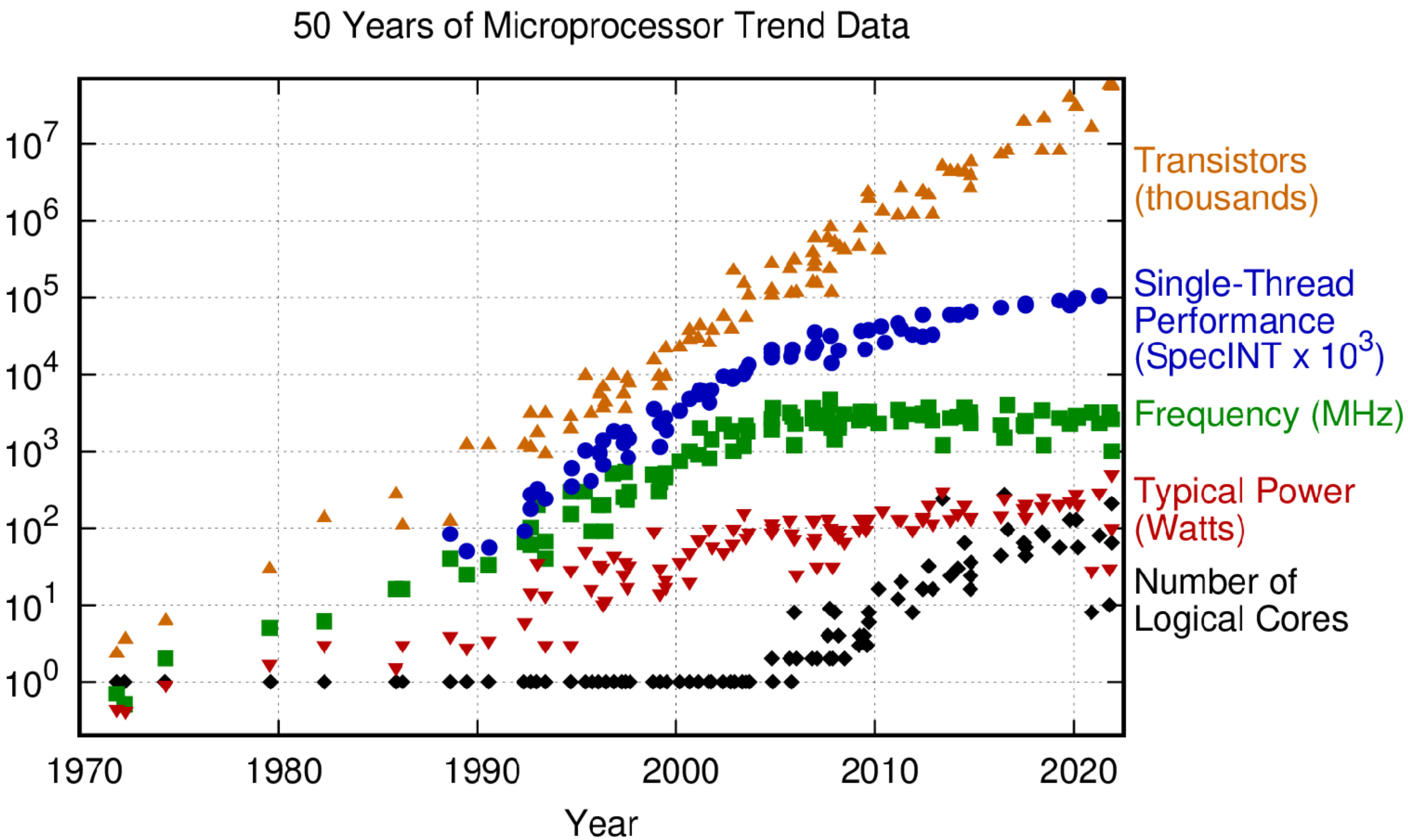
> "Current practice for large general-purpose experiments … is to use at least two more entities, colloquially referred to as the "Level-2" and "Level-3" trigger."

⇒ Decision based on Moore's law scaling.

* Advanced computing ca. 2000
= more (complex) compute on faster processors

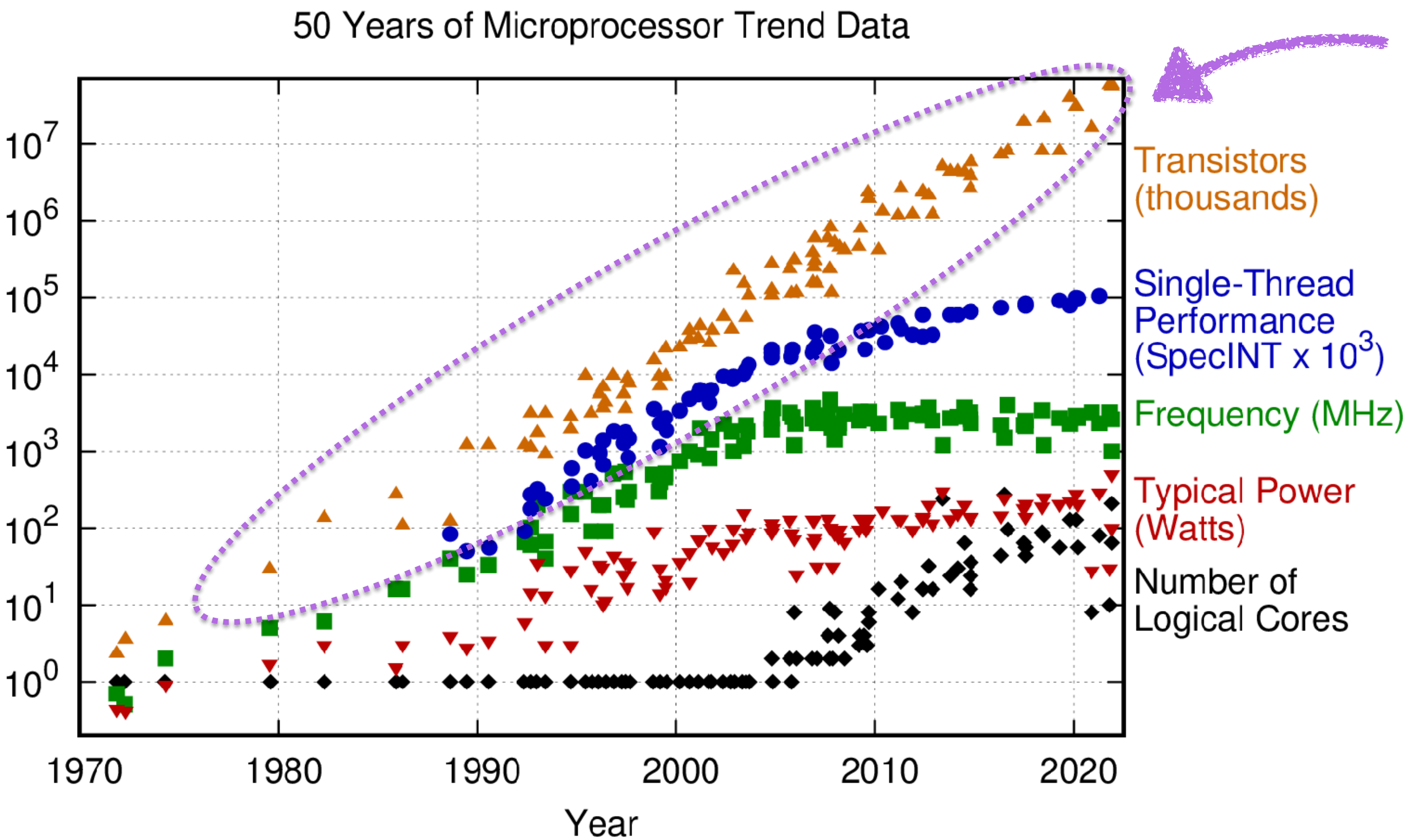
Computing technology evolves fast; be nimble & bold

Processors are not getting faster



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

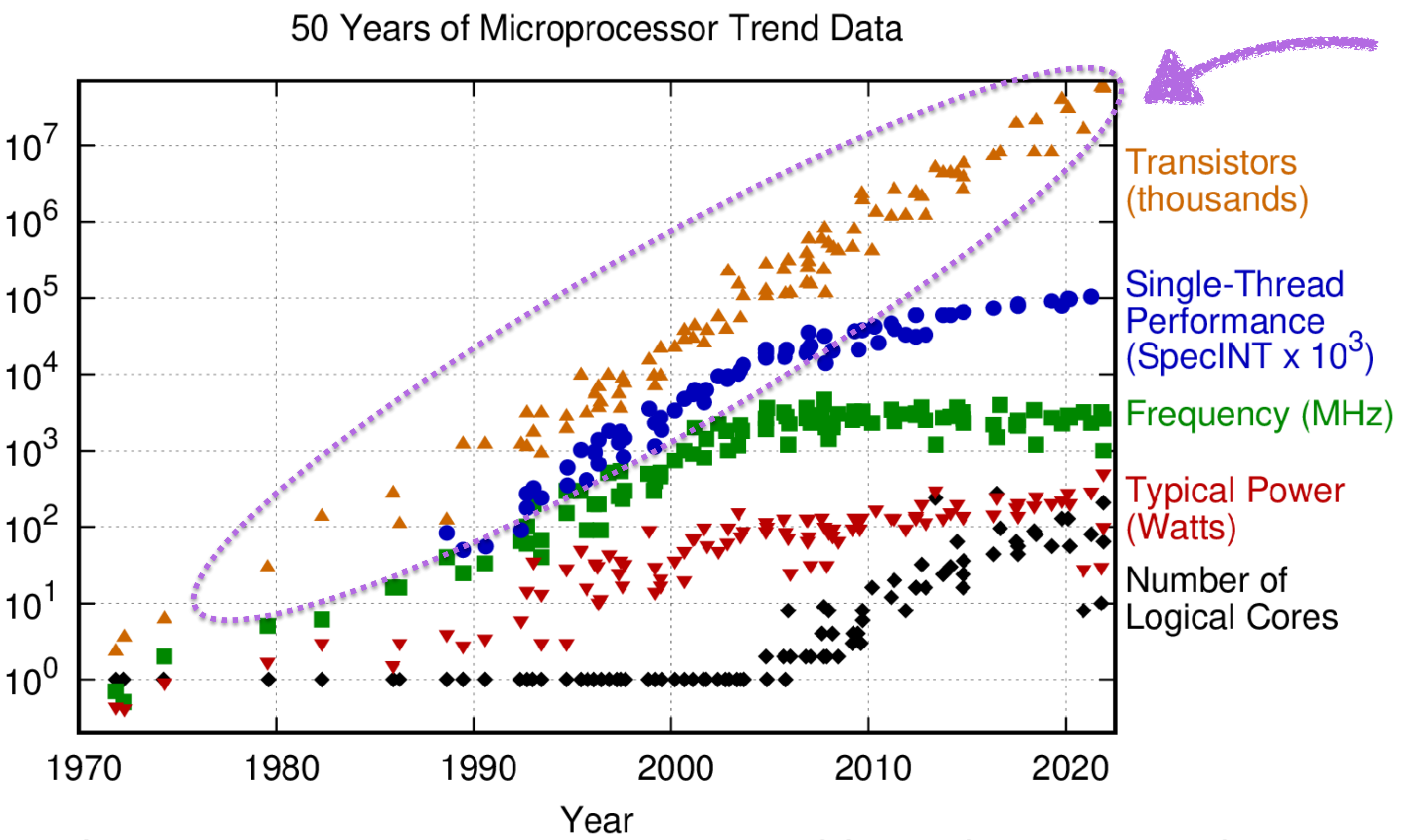
Processors are not getting faster



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

Moore's law still holds for now

Processors are not getting faster

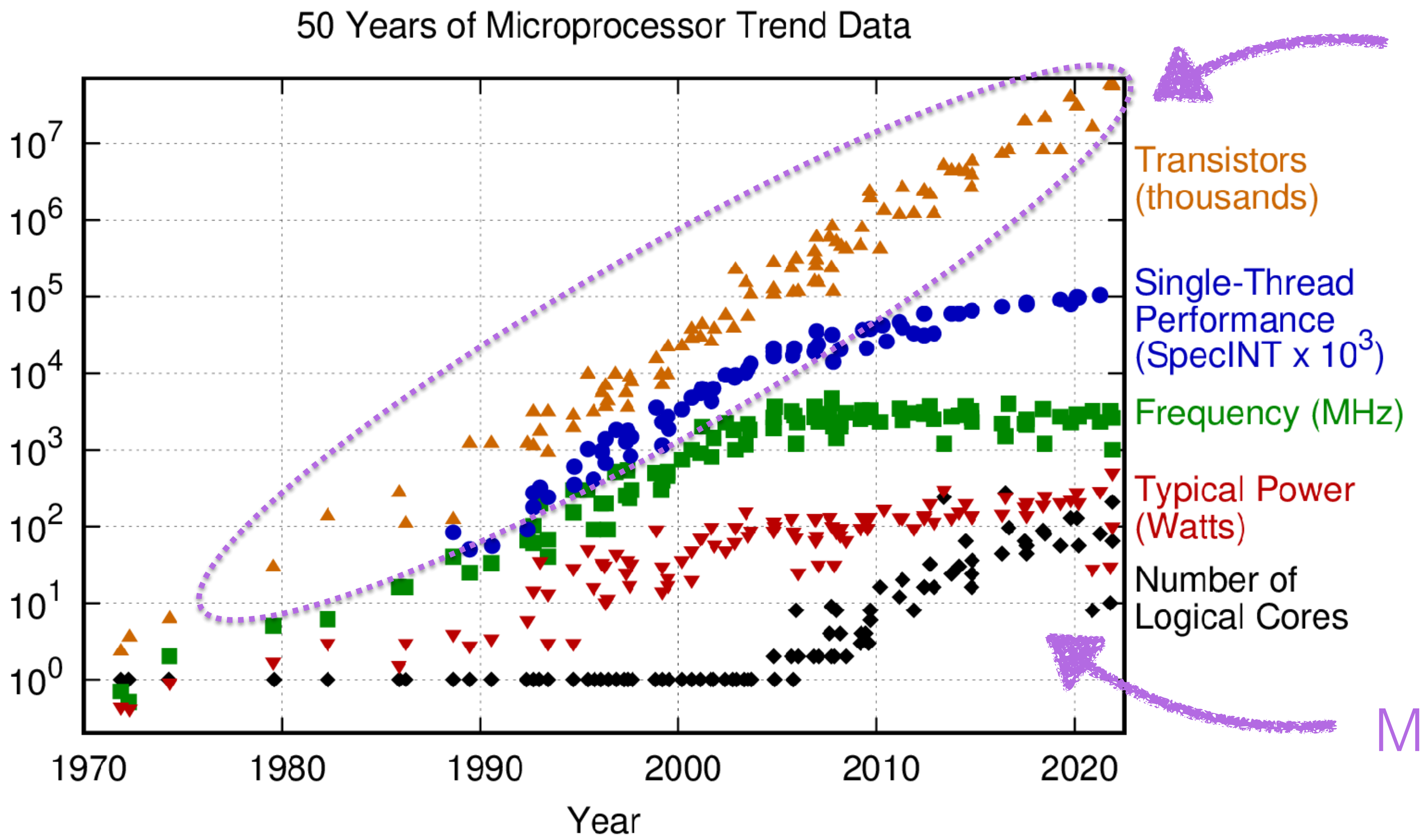


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

Moore's law still holds for now

But clock frequency has been flat for 15 years

Processors are not getting faster



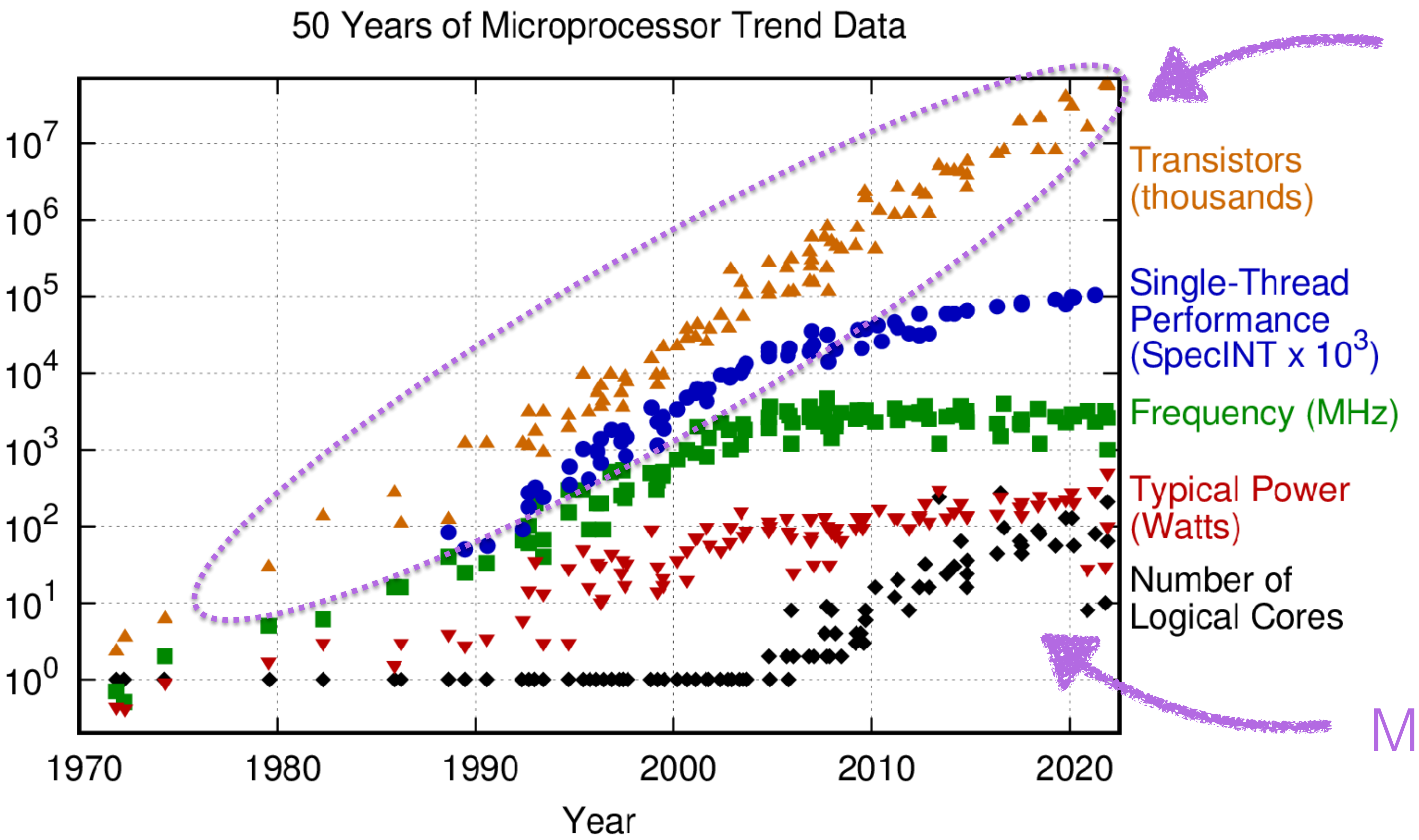
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

Moore's law still holds for now

But clock frequency has been flat for 15 years

Many-core has become the norm

Processors are not getting faster



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

Moore's law still holds for now

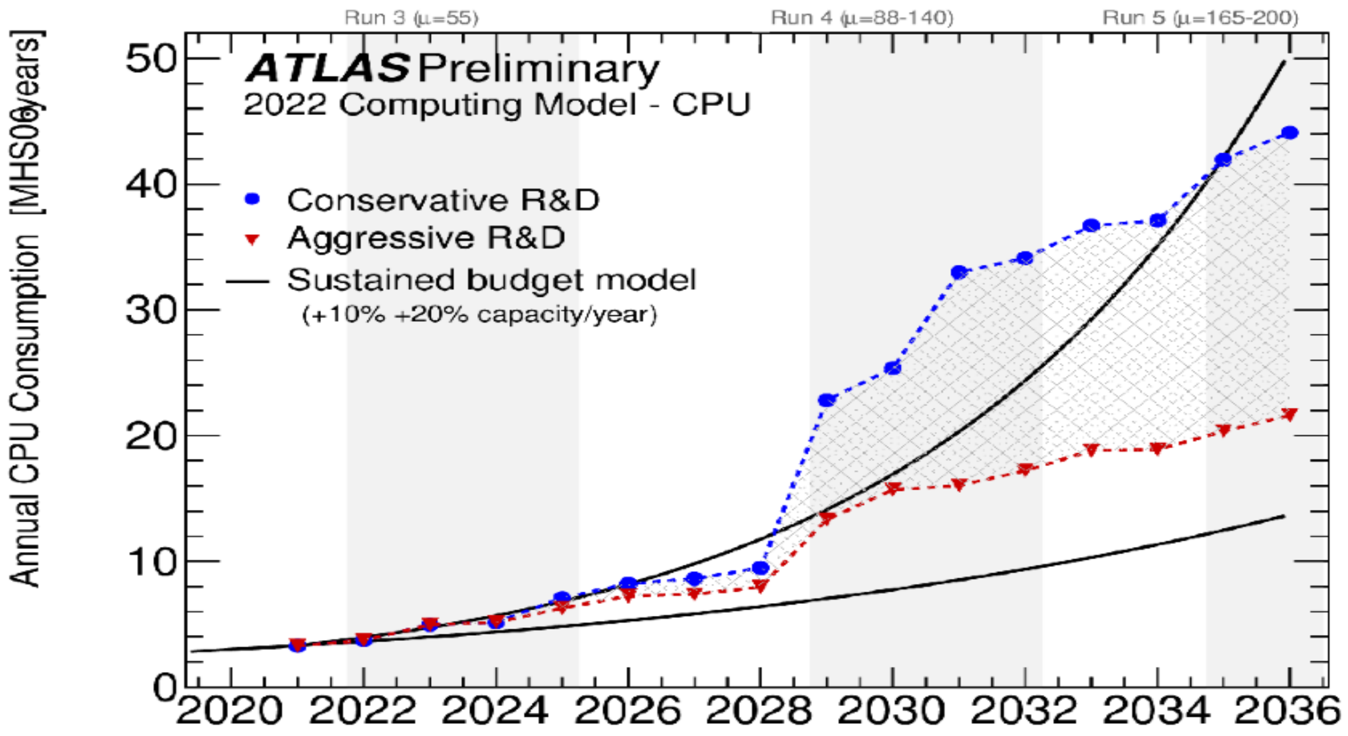
But clock frequency has been flat for 15 years

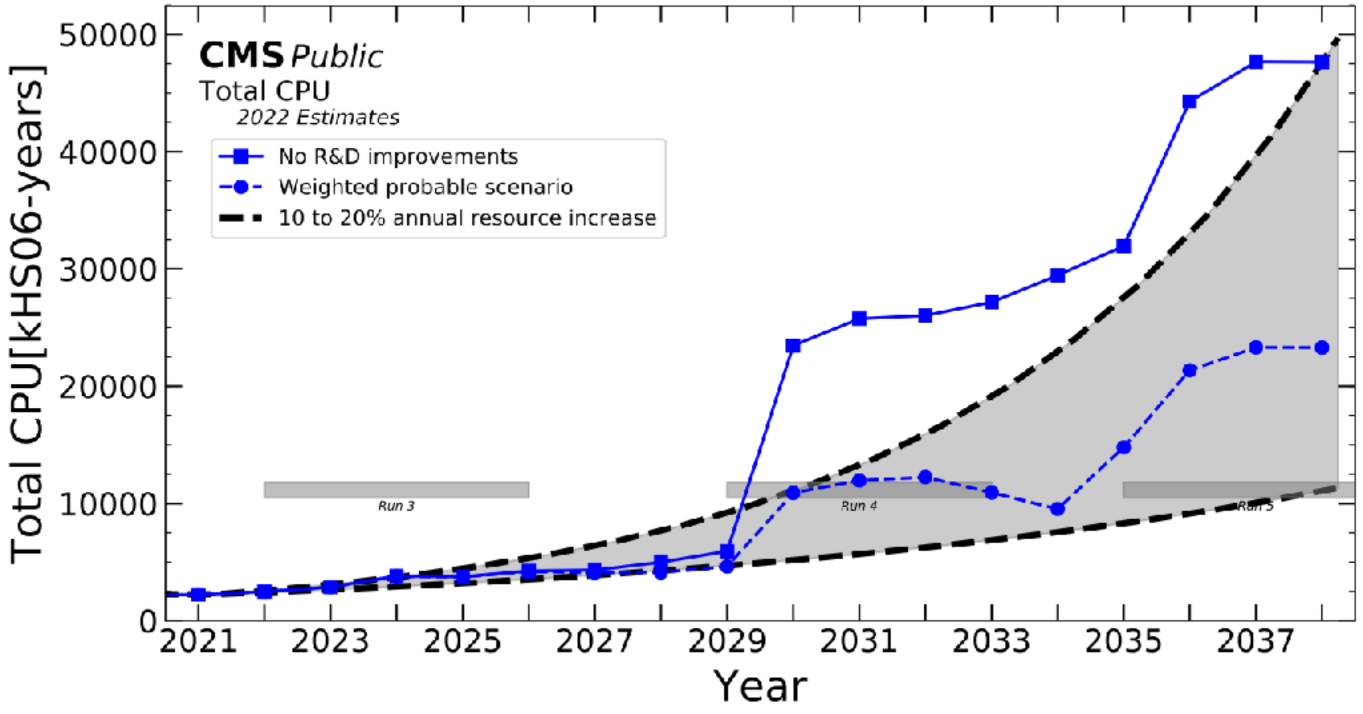
Many-core has become the norm

“Advanced computing” means something different now

HEP implications - absolute needs

- CPU need jumps up in the HL-LHC era
  - More recorded data
    → More reconstruction
    → More simulation
  - Higher order simulation
- Capacity scaling not sufficient
- Computing techniques must be updated





HEP implications - opportunities

- Machine learning - for all sorts of applications
- Acceleration with GPU etc.
  - Faster analysis (event selection, fitting, ...)
  - Higher trigger bandwidth
- On- / near-detector intelligence
- Efficient and accurate simulation

Many interesting ideas reported at CHEP 2023

Technological enablers

Graphical processing units

What they are:

Many (O(1000)) processor cores with simple instructions + fast memory

→ Optimized for ~identical operations on elements of large arrays

Applications:

- Neural network evaluation → deep learning
- Cellular automaton → track seed finding, clustering
- Fitting
- Any parallel calculation over arrays

References: 2008.13461 Rembser, CHEP2023

AI ASICs and other specialized chips

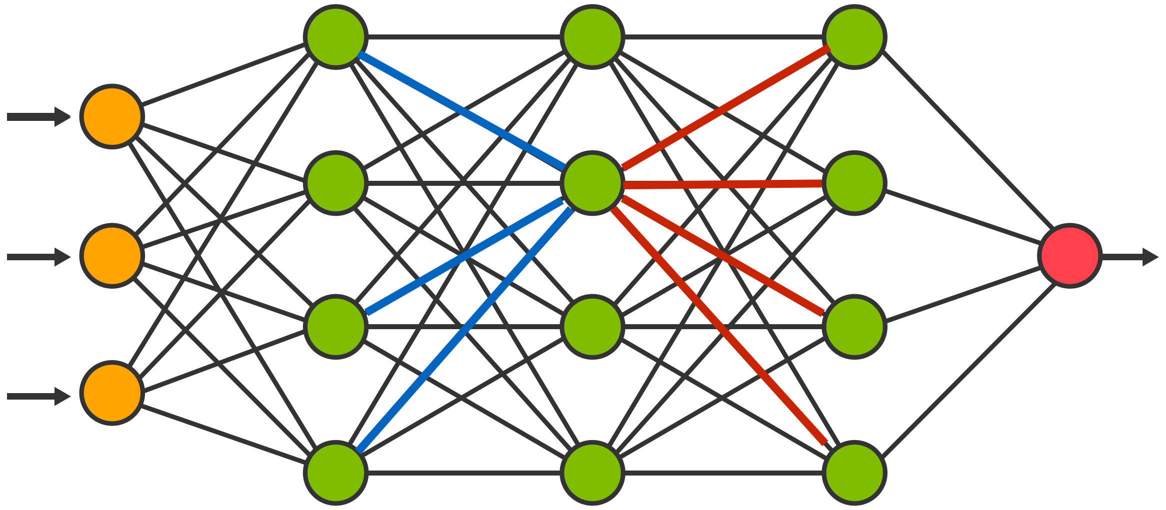
TPUs, IPUs, ...: Different optimizations of parallel processors

- Reduced numeric precision
- Hard-wired matrix operation
- etc.

Edge TPU: small form factor + low power consumption

Neuromorphic processors

- Artificial neuron fires asynchronously only when input sum exceeds threshold
  → Low power, low latency
- Neurons updated on the fly → adaptive



High-performance computing

What it is:

A cluster of high-spec (often custom-built) compute nodes with fast interconnect

→ Makes a cluster look like a single very-many-core computer

Applications:

- Multiprocess tasks
  - Recent HPCs tend to be GPU clusters
    → Large-scale ML training
- HEP tasks not best suited, but investments in compute are increasingly concentrated on HPCs
  → Need to learn to use them

Machine learning

- *The* biggest advancement of the past decade
- Must be mentioned here, but already covered by Matt
- Large language models?

Differentiable programming

Two enablers of deep learning:

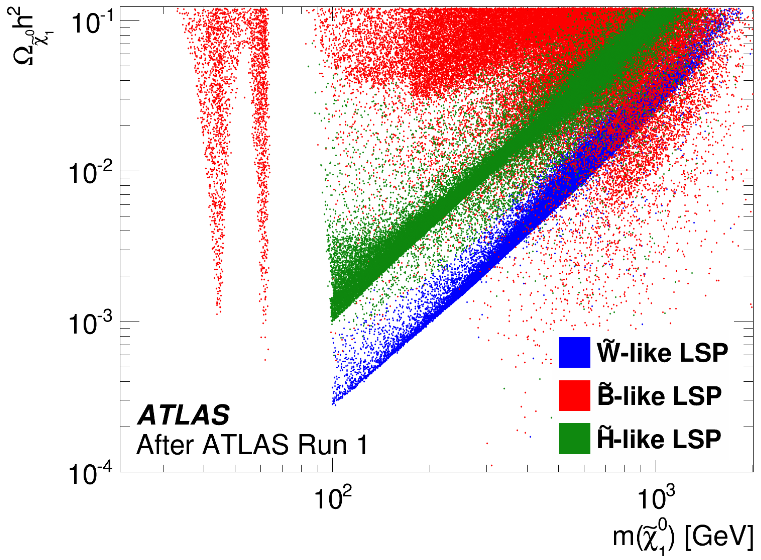
- Parallel computation
- Gradient-descent optimization
  → Powered by auto-differentiation

```
def func(x):
    return x * x + 3 * x

func(2) # = 10
grad(func)(2) # 2x+3 = 7
```

What can we calculate with derivatives?

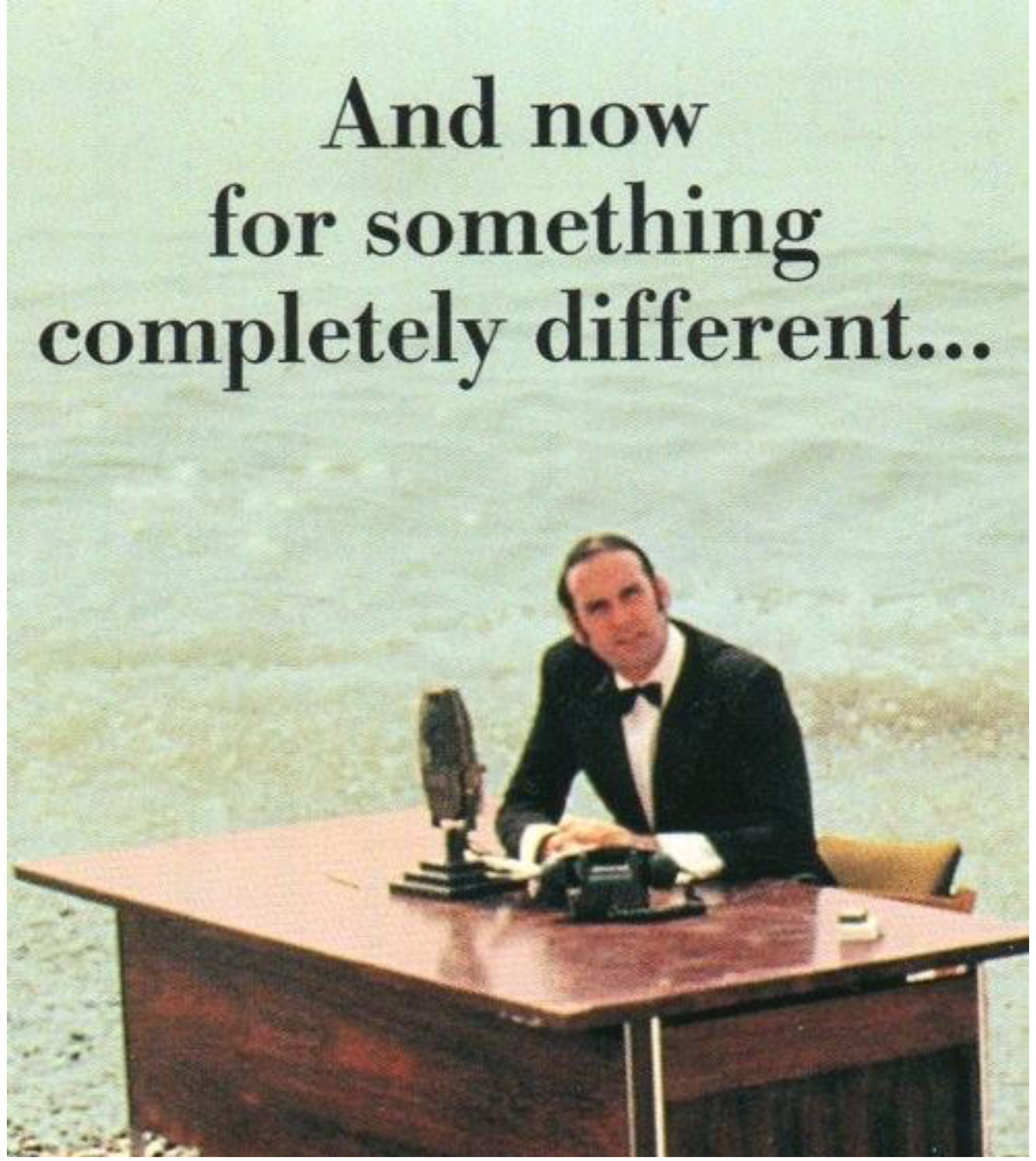
- Analysis sensitivity as a function of UV model parameters
- Reconstruction efficiency as a function of detector geometry
- Limits / significances as a function of calibration constants
  event selection
  …



No more need for toy MCs?

Refs: 2203.00057 https://mode-collaboration.github.io/ G. Singh, CHEP2023

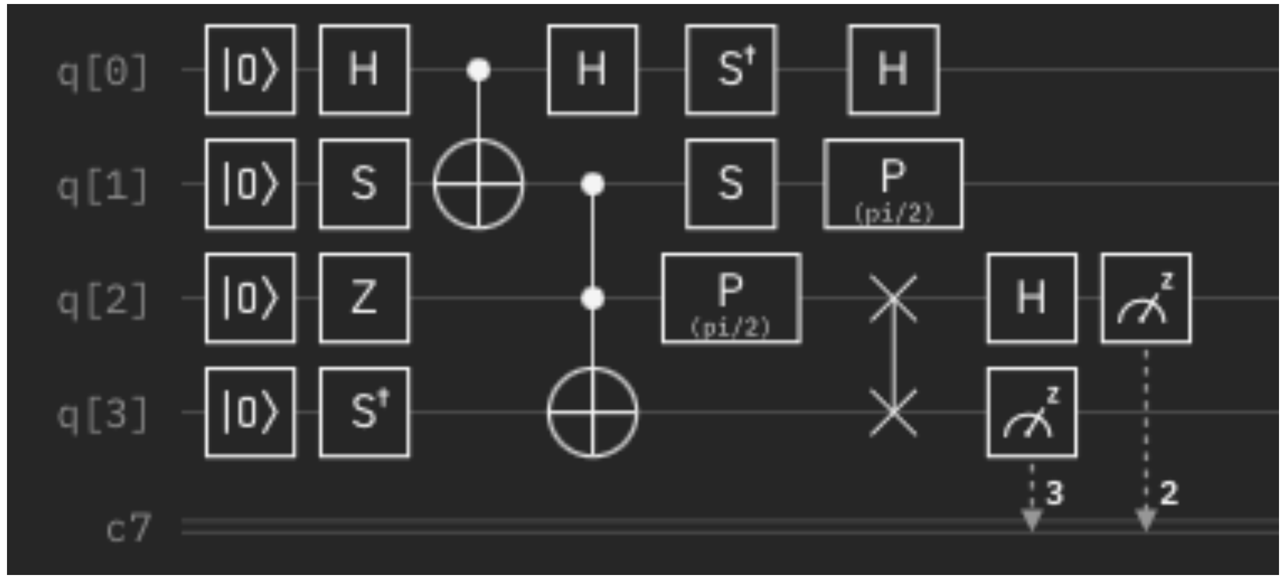
Quantum computation



Introduction

“Quantum circuit model” computer:

- Comprises n two-level quantum systems
    - Two levels labeled as $|0\rangle$ and $|1\rangle$
- Can prepare the systems collectively into an initial state $|0\rangle^{\otimes n}$
- Can apply well-defined unitary operations on systems in arbitrary sequences
- Allows projective measurement of system states in the $\{|0\rangle, |1\rangle\}$ basis

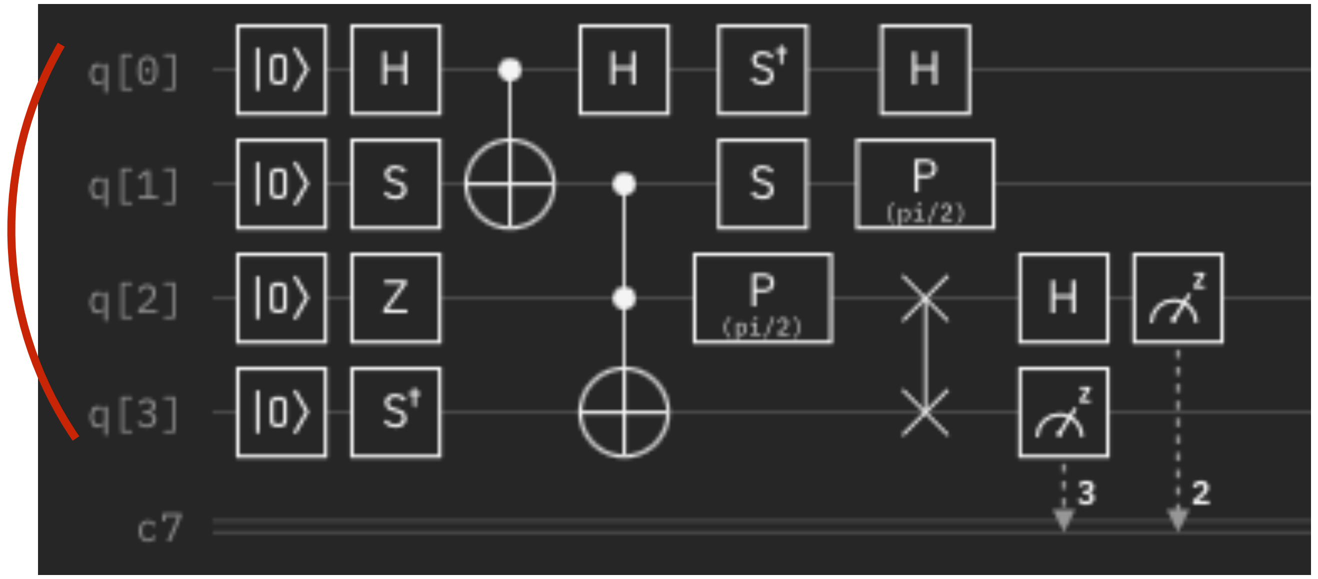


Introduction

"Quantum circuit model" computer:

qubits

- Comprises n two-level quantum systems
  - Two levels labeled as $|0\rangle$ and $|1\rangle$
- Can prepare the systems collectively into an initial state $|0\rangle^{\otimes n}$
- Can apply well-defined unitary operations on systems in arbitrary sequences
- Allows projective measurement of system states in the $\{|0\rangle, |1\rangle\}$ basis



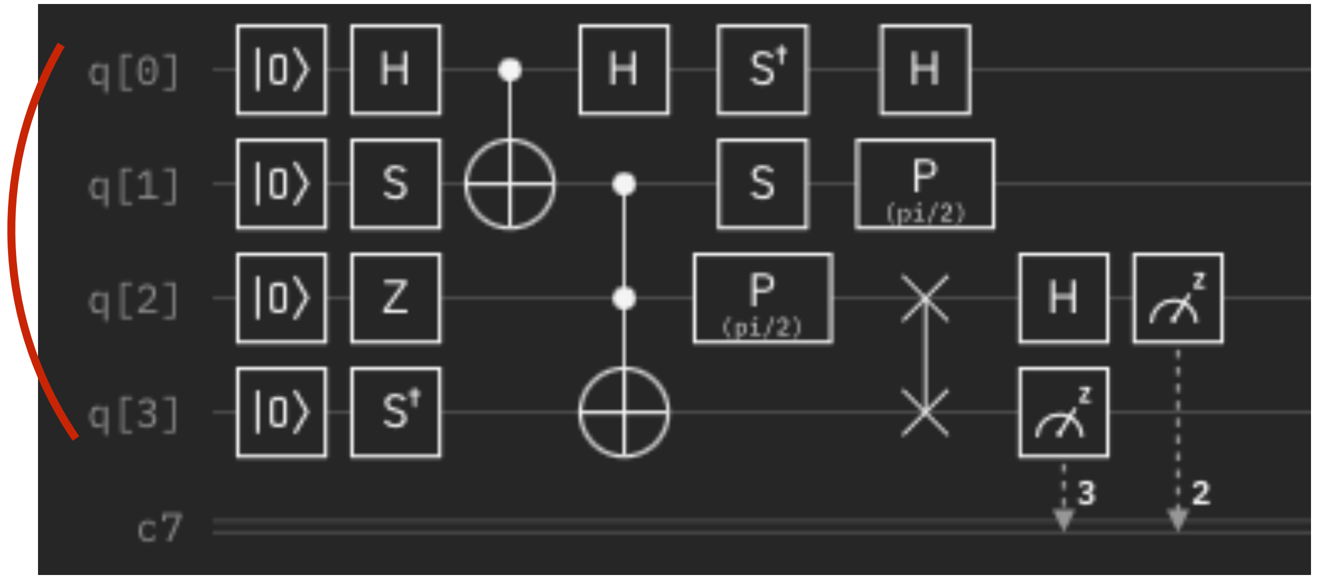
Introduction

“Quantum circuit model” computer:

- Comprises n two-level quantum systems
  - Two levels labeled as $|0\rangle$ and $|1\rangle$
- Can prepare the systems collectively into an initial state $|0\rangle^{\otimes n}$
- Can apply well-defined unitary operations on systems in arbitrary sequences
- Allows projective measurement of system states in the $\{|0\rangle, |1\rangle\}$ basis

qubits

quantum gates



Introduction

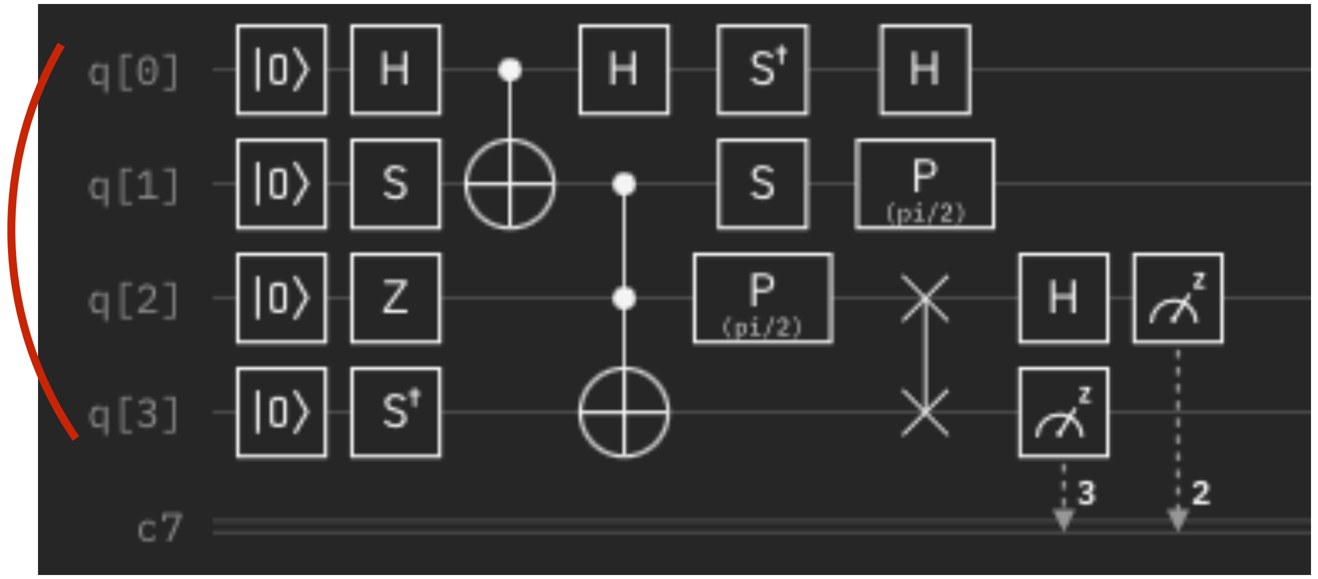
“Quantum circuit model” computer:

- Comprises n two-level quantum systems
  - Two levels labeled as $|0\rangle$ and $|1\rangle$
- Can prepare the systems collectively into an initial state $|0\rangle^{\otimes n}$
- Can apply well-defined unitary operations on systems in arbitrary sequences
- Allows projective measurement of system states in the $\{|0\rangle, |1\rangle\}$ basis

qubits

quantum gates

quantum circuits



Introduction

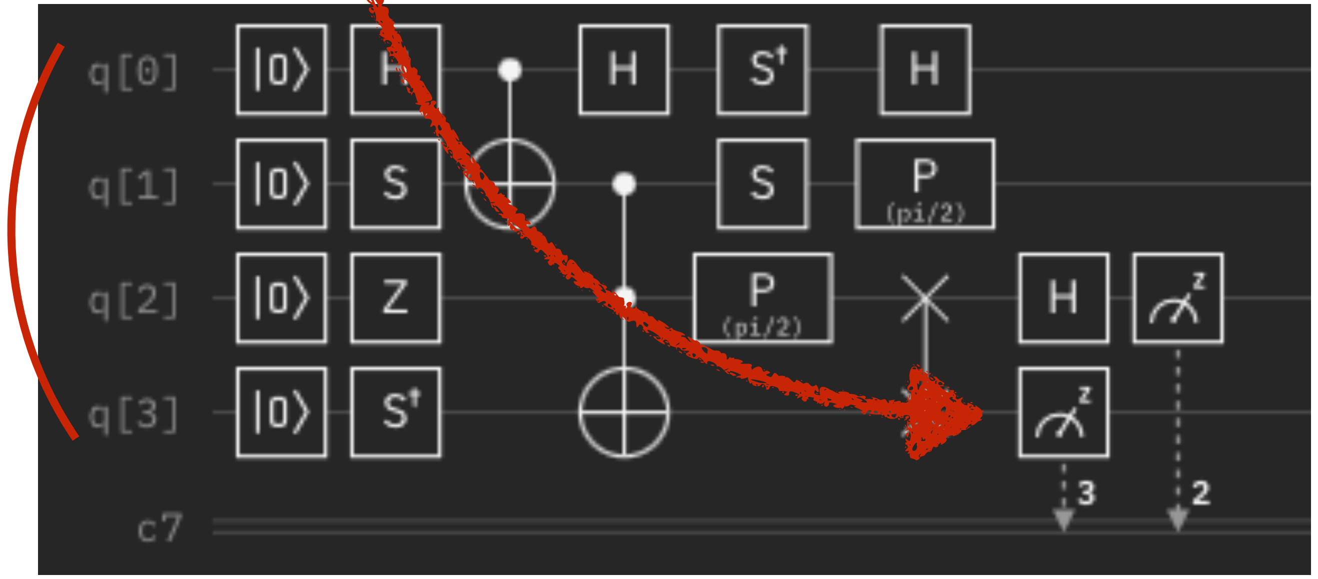
"Quantum circuit model" computer:

- Comprises n two-level quantum systems
    - Two levels labeled as $|0\rangle$ and $|1\rangle$
- Can prepare the systems collectively into an initial state $|0\rangle^{\otimes n}$
- Can apply well-defined unitary operations on systems in arbitrary sequences
- Allows projective measurement of system states in the $\{|0\rangle, |1\rangle\}$ basis

qubits

quantum gates

quantum circuits



Exponential dimensions

- n-qubit Hilbert space: 2^n-dimensional

$$|\psi\rangle = \psi_0|00\ldots00\rangle + \psi_1|00\ldots01\rangle + \cdots + \psi_{2^n-1}|11\ldots11\rangle$$

- *In principle* we can create arbitrary states with gate operations.

→ A computer with $2^{\mathcal{O}(100)}$ parameters!

cf. Number of atoms in the observable universe ~ 2^{270}

Exponential dimensions

- n-qubit Hilbert space: 2^n-dimensional

$$|\psi\rangle = \psi_0|00\ldots00\rangle + \psi_1|00\ldots01\rangle + \cdots + \psi_{2^n-1}|11\ldots11\rangle$$

- *In principle* we can create arbitrary states with gate operations.

→ A computer with $2^{\mathcal{O}(100)}$ parameters!

cf. Number of atoms in the observable universe ~ 2^{270}

Not quite.

Exponential dimensions

- n-qubit Hilbert space: 2^n-dimensional

$$|\psi\rangle = \psi_0|00\ldots00\rangle + \psi_1|00\ldots01\rangle + \cdots + \psi_{2^n-1}|11\ldots11\rangle$$

- *In principle* we can create arbitrary states with gate operations.

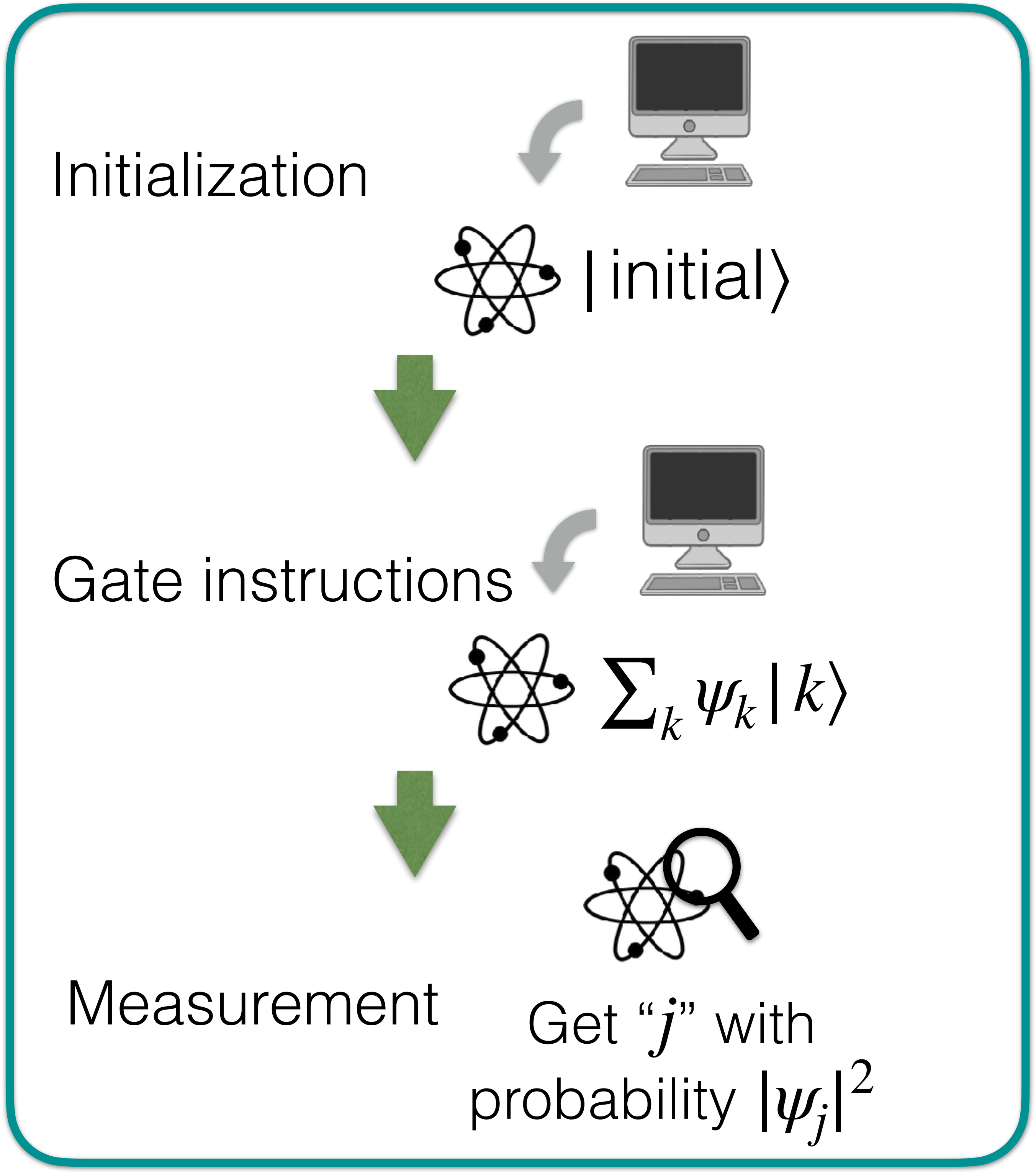
→ A computer with $2^{\mathcal{O}(100)}$ parameters!
cf. Number of atoms in the observable universe ~ 2^{270}

Not quite.

- Need exponentially many gates to manipulate the qubits over full space
- We can only measure $|\psi_j|^2$

→ “Quantum advantage” only for specific classes of problems

Basic usage



Initialization

$|\text{initial}\rangle$

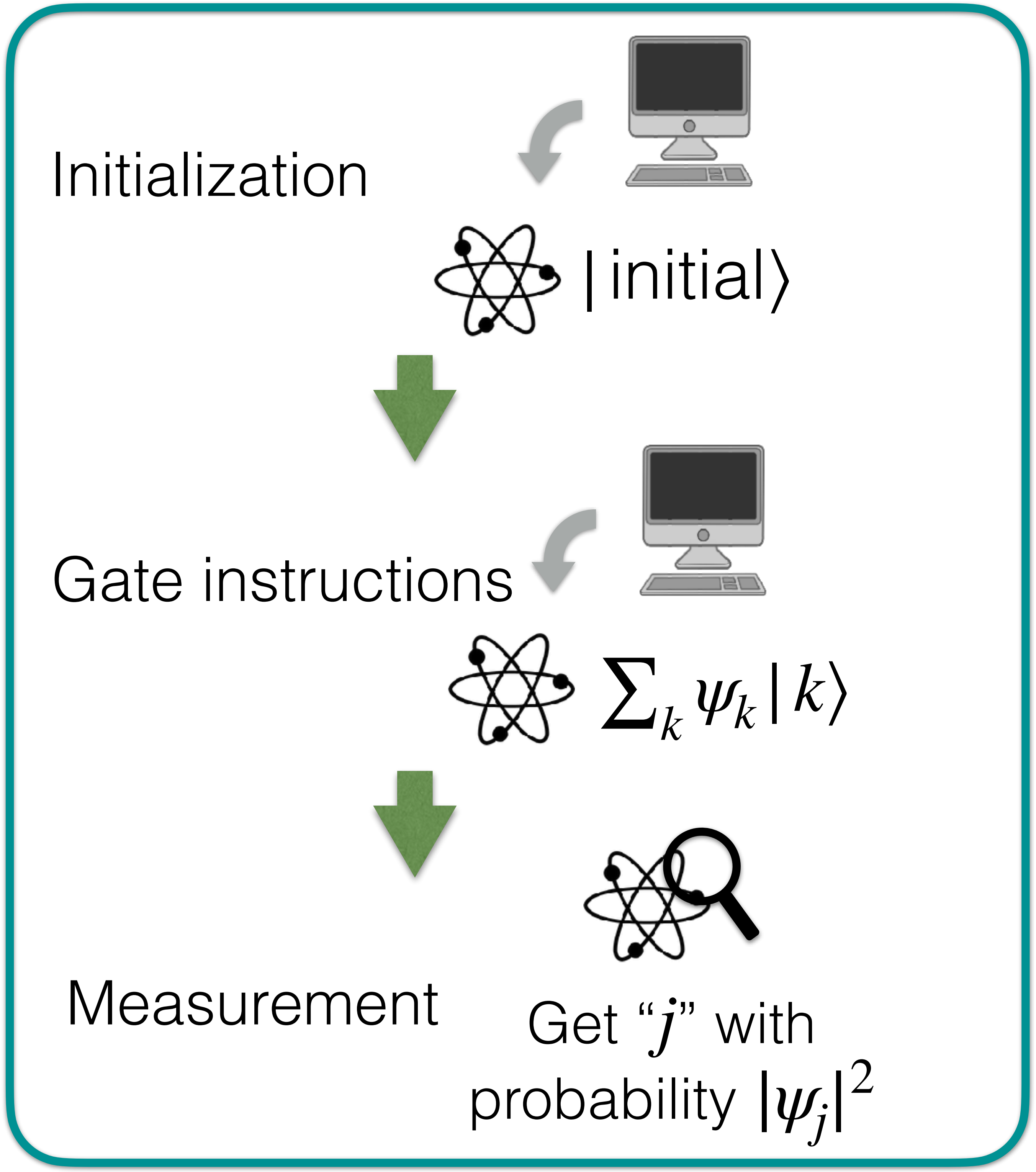
Gate instructions

$\sum_k \psi_k |k\rangle$

Measurement

Get “j” with probability $|\psi_j|^2$

Basic usage



Initialization

$|\text{initial}\rangle$

Gate instructions

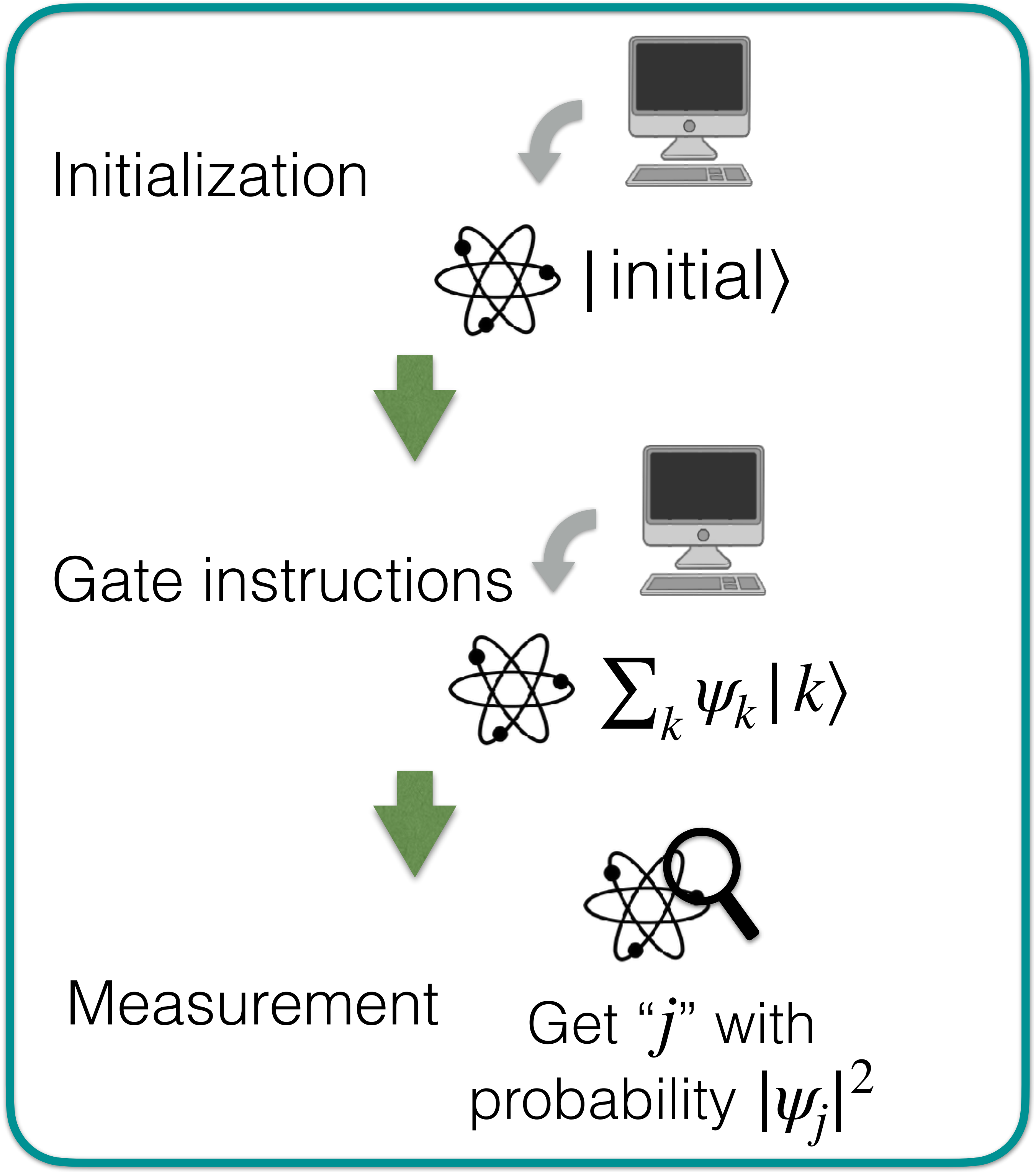
$\sum_k \psi_k |k\rangle$

Measurement

Get “j” with probability $|\psi_j|^2$

× Many trials

Basic usage



Initialization

$|\mathrm{initial}\rangle$

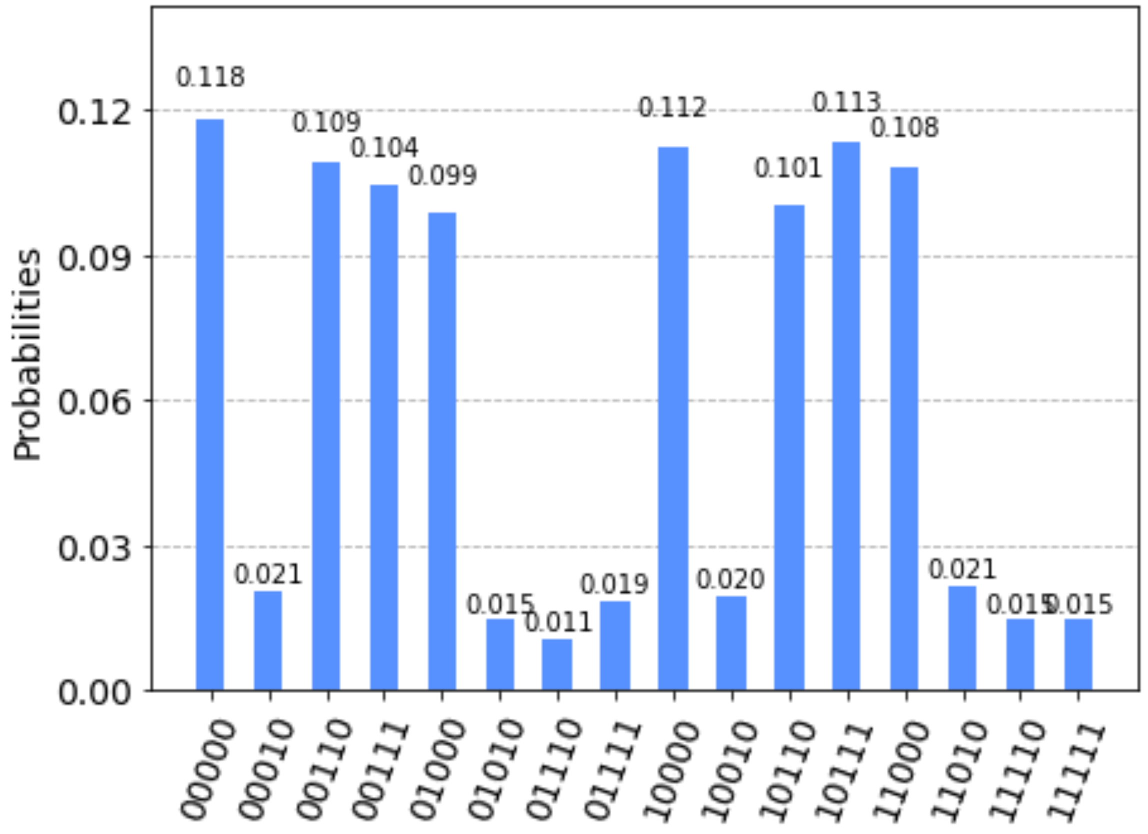
Gate instructions

$\sum_k \psi_k |k\rangle$

Measurement

Get "j" with probability $|\psi_j|^2$

× Many trials



Analyze the distribution

- Expectation value
- Mode
- etc.

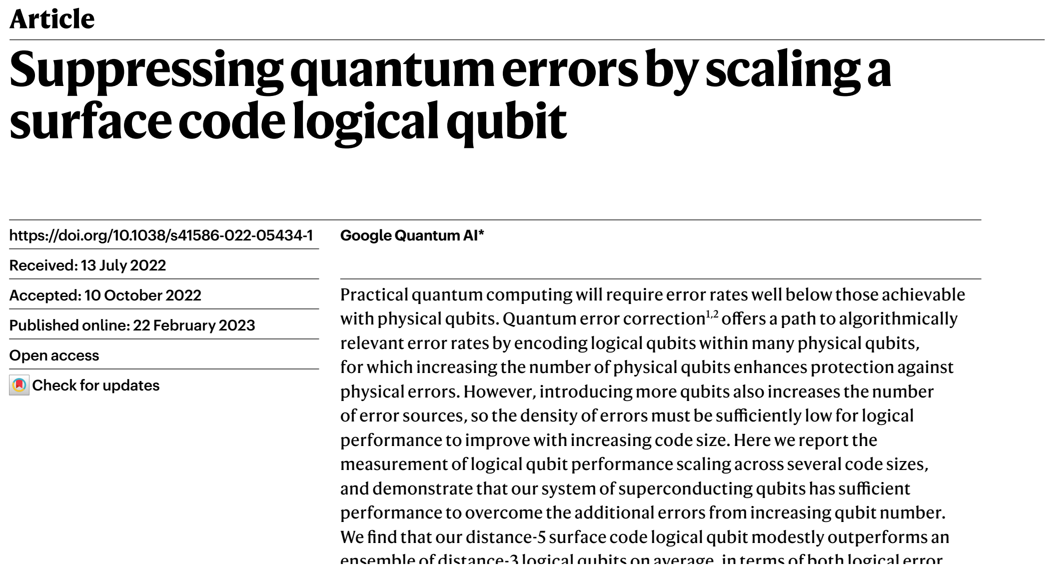
Error correction and mitigation

- Qubits must be controllable → Not an isolated system → Decoherence
- Gate operations = analog control

⇒ Lots of errors occur at device level

Real-time error correction required for realistic-depth circuits

Error correction is coming:



Article

Suppressing quantum errors by scaling a surface code logical qubit

https://doi.org/10.1038/s41586-022-05434-1

Received: 13 July 2022

Accepted: 10 October 2022

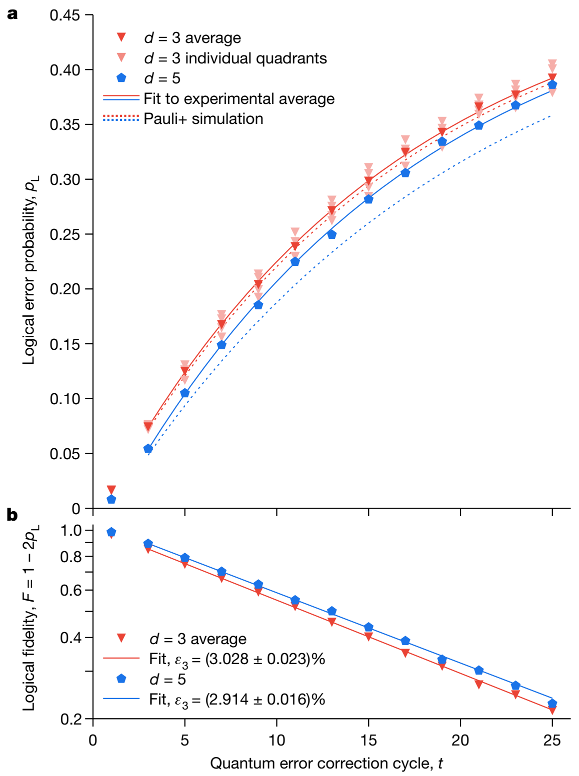
Published online: 22 February 2023

Open access

Check for updates

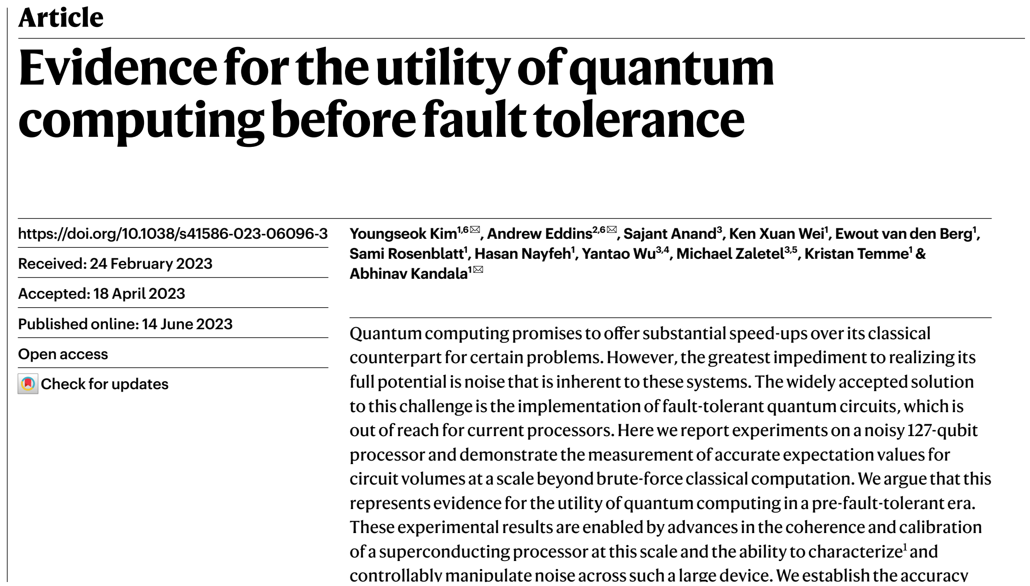
Google Quantum AI*

Practical quantum computing will require error rates well below those achievable with physical qubits. Quantum error correction[1,2] offers a path to algorithmically relevant error rates by encoding logical qubits within many physical qubits, for which increasing the number of physical qubits enhances protection against physical errors. However, introducing more qubits also increases the number of error sources, so the density of errors must be sufficiently low for logical performance to improve with increasing code size. Here we report the measurement of logical qubit performance scaling across several code sizes, and demonstrate that our system of superconducting qubits has sufficient performance to overcome the additional errors from increasing qubit number. We find that our distance-5 surface code logical qubit modestly outperforms an ensemble of distance-3 logical qubits on average, in terms of both logical error



Google AI, Nature 614, 676 (2023)

But also:



Article

Evidence for the utility of quantum computing before fault tolerance

https://doi.org/10.1038/s41586-023-06096-3

Received: 24 February 2023

Accepted: 18 April 2023

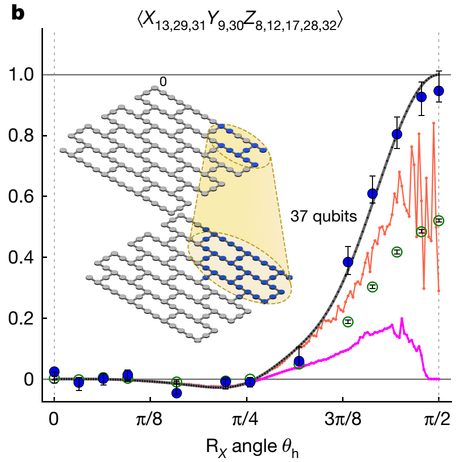
Published online: 14 June 2023

Open access

Check for updates

Youngseok Kim[1,6] ✉, Andrew Eddins[2,6] ✉, Sajant Anand[3], Ken Xuan Wei[1], Ewout van den Berg[1], Sami Rosenblatt[1], Hasan Nayfeh[1], Yantao Wu[3,4], Michael Zaletel[3,5], Kristan Temme[1] & Abhinav Kandala[1] ✉

Quantum computing promises to offer substantial speed-ups over its classical counterpart for certain problems. However, the greatest impediment to realizing its full potential is noise that is inherent to these systems. The widely accepted solution to this challenge is the implementation of fault-tolerant quantum circuits, which is out of reach for current processors. Here we report experiments on a noisy 127-qubit processor and demonstrate the measurement of accurate expectation values for circuit volumes at a scale beyond brute-force classical computation. We argue that this represents evidence for the utility of quantum computing in a pre-fault-tolerant era. These experimental results are enabled by advances in the coherence and calibration of a superconducting processor at this scale and the ability to characterize[1] and controllably manipulate noise across such a large device. We establish the accuracy



Kim et al., Nature 618, 500 (2023)

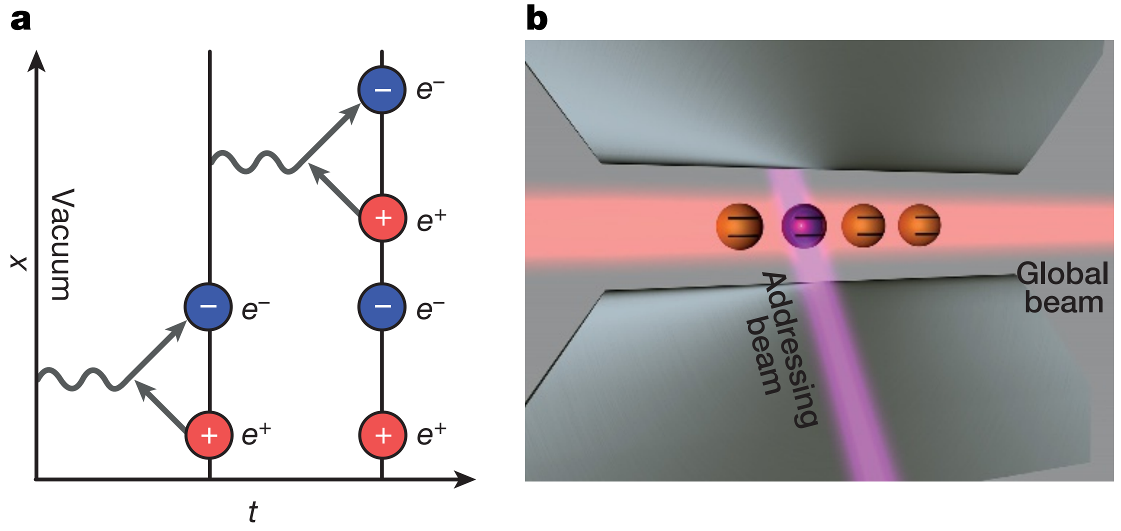
Application: Lattice QFT simulation

Simple recipe:

Jordan et al. Science 336, 1130 (2012)

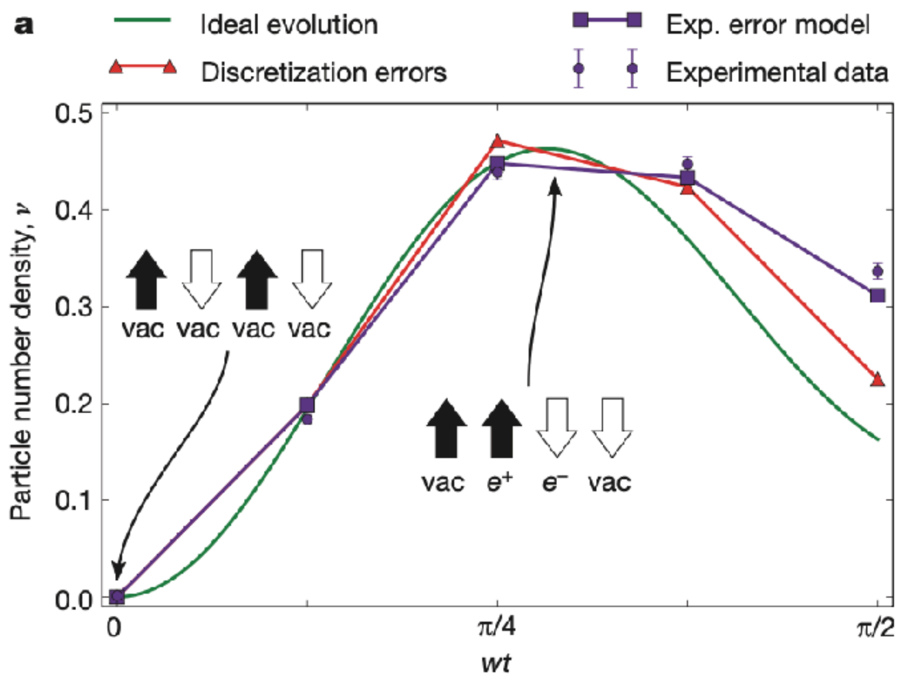
1. Discretize space
2. Represent field at each node / link with qubits
3. Prepare initial state
4. Express time evolution with gates
5. Simulate and measure

→ Full dynamics simulation of quantum fields!



Schwinger model (1+1D QED) demonstrated in 2016 on a 4-bit trapped ion system

Martinez et al. Nature 534, 516 (2016)



Simple recipe is insufficient

In practice:

- Long simulation
  → need error correction
- Requires many qubits

Energy range expressible by lattice:

$$\frac{1}{Nl} < E < \frac{1}{l}$$

+

Full simulation of LHC scattering:

$$100\mathrm{MeV} < E < 7\mathrm{TeV}$$

Number of lattice sites

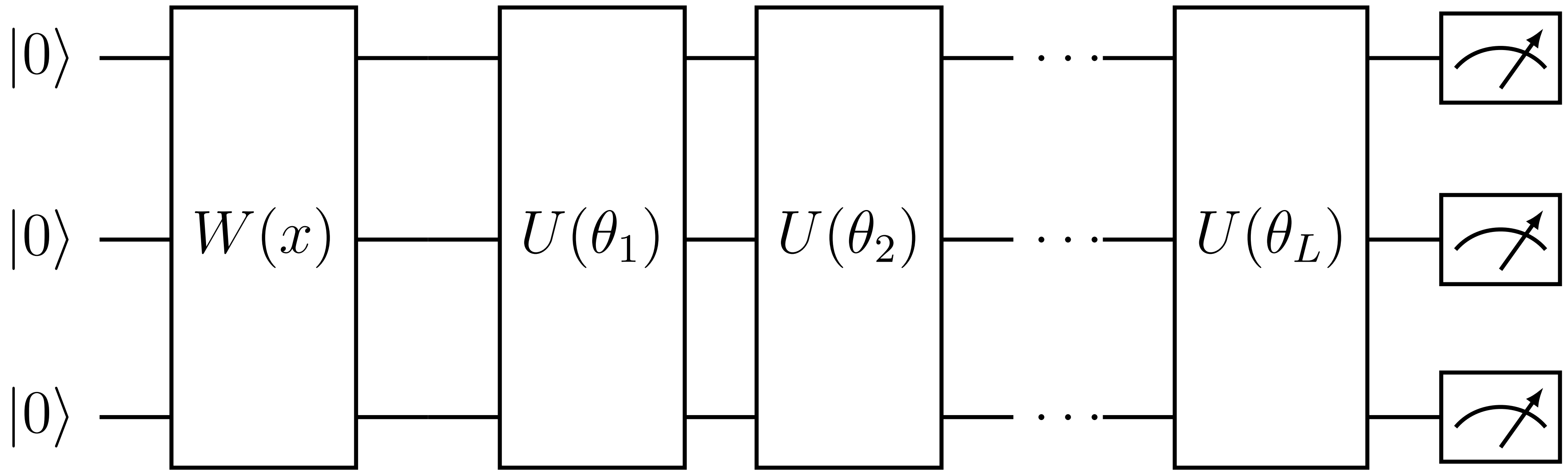
$$N^3 \sim 10^{14}$$

Ideas to lower qubit requirement:

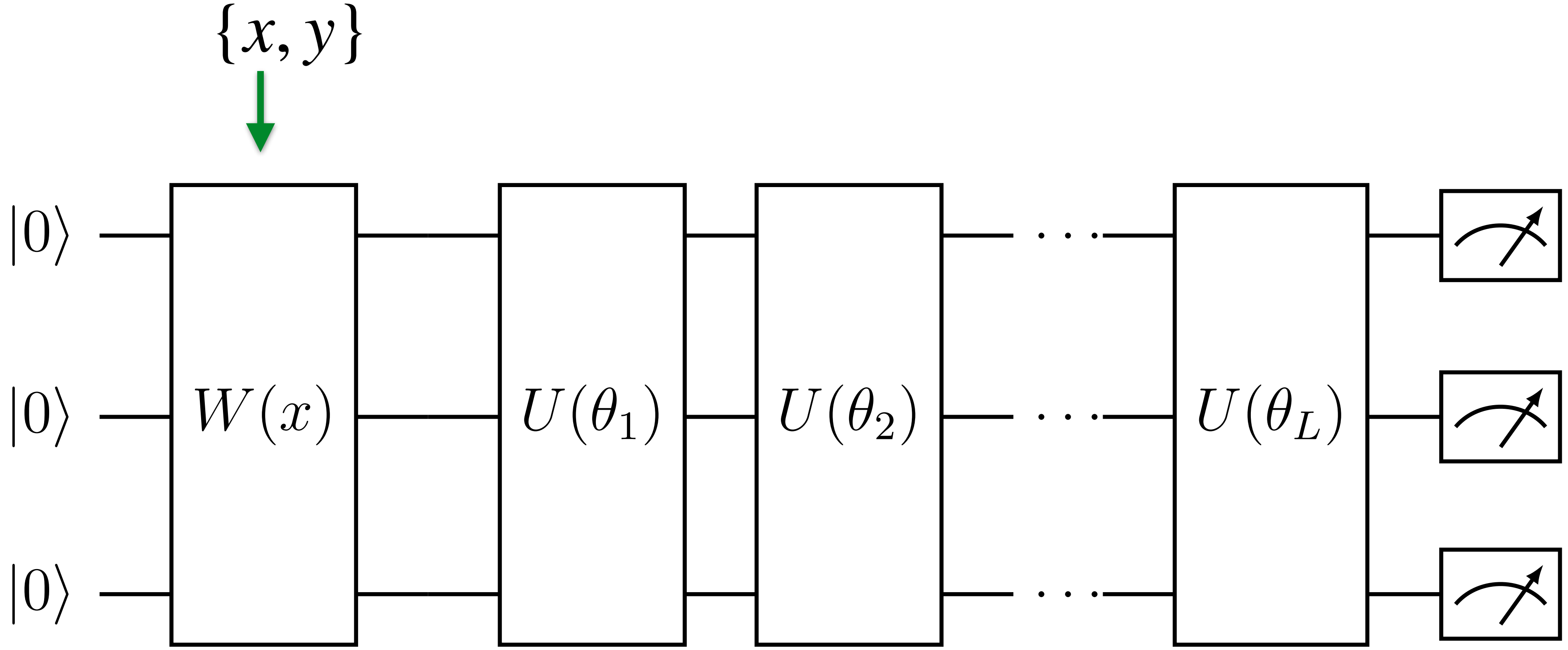
- EFT: division of phase space (2102.05044)
- Particle-based field representation (2012.00020)

Application: Quantum “neural network”

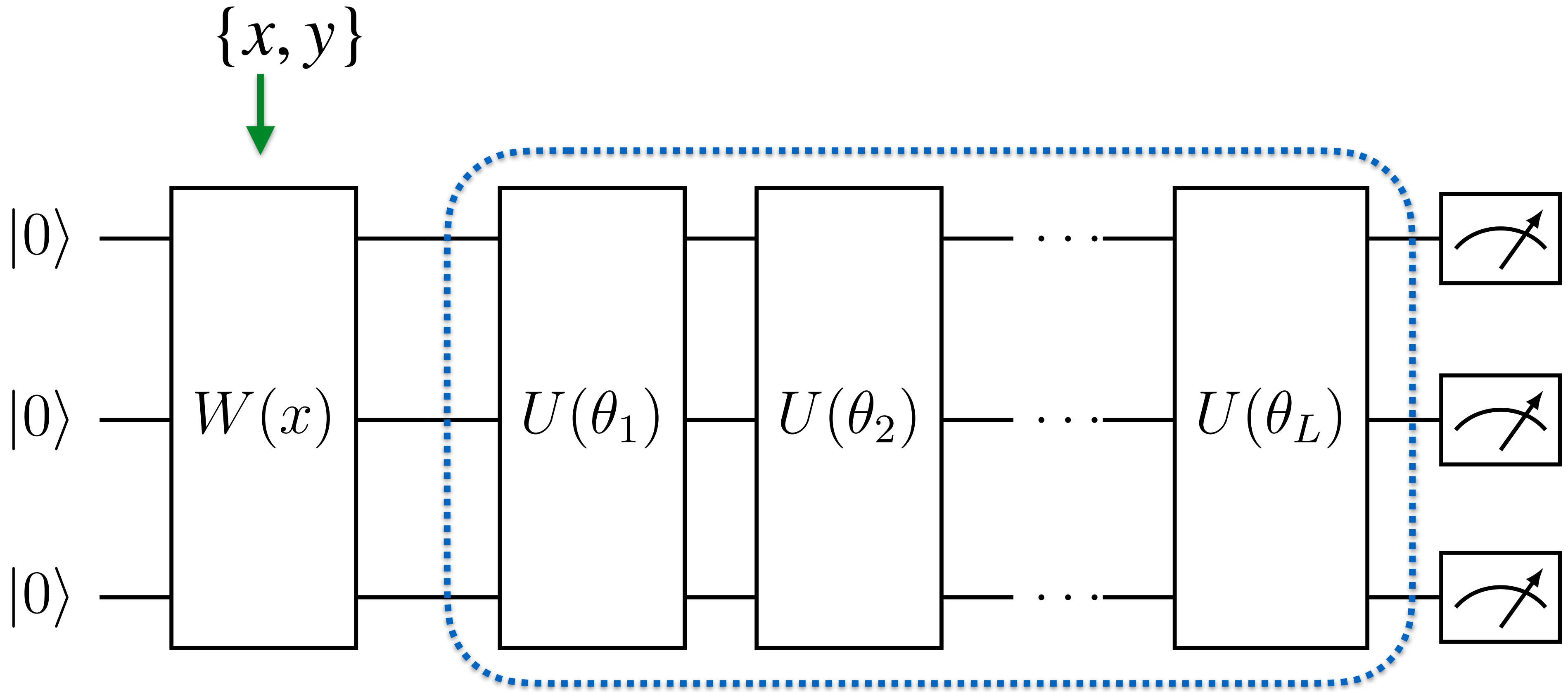
$\{x, y\}$



Application: Quantum “neural network”

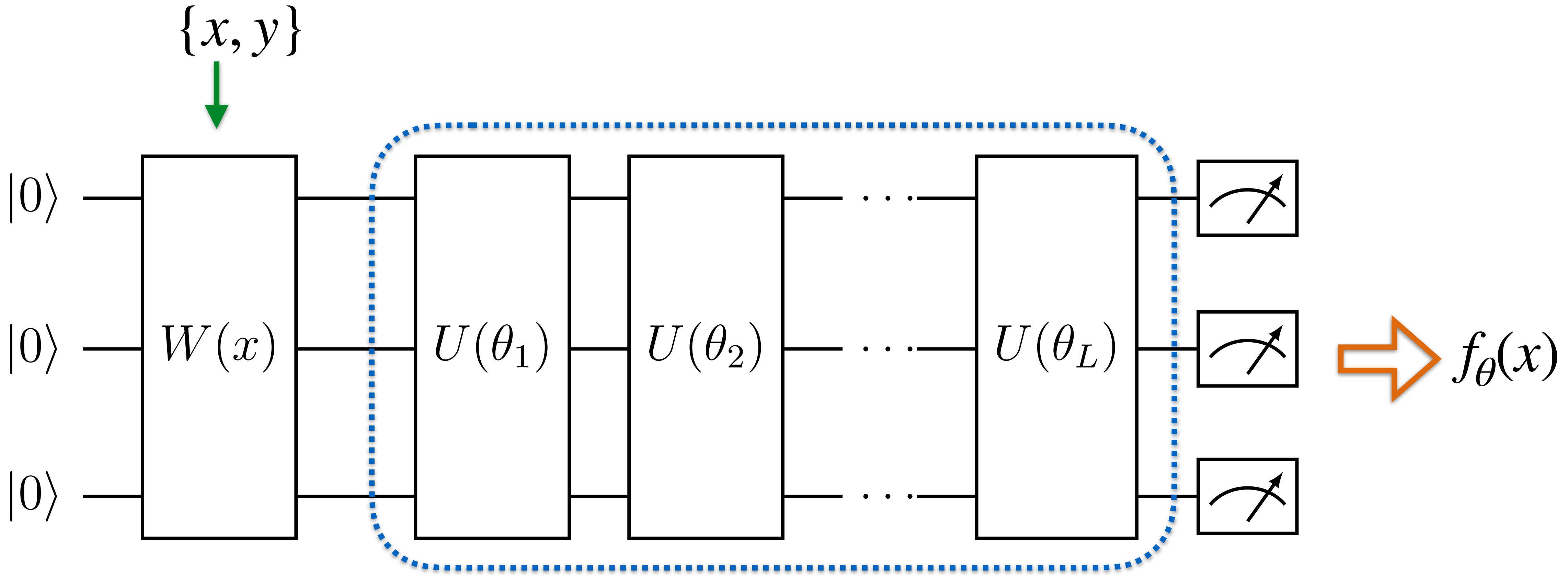


Application: Quantum “neural network”



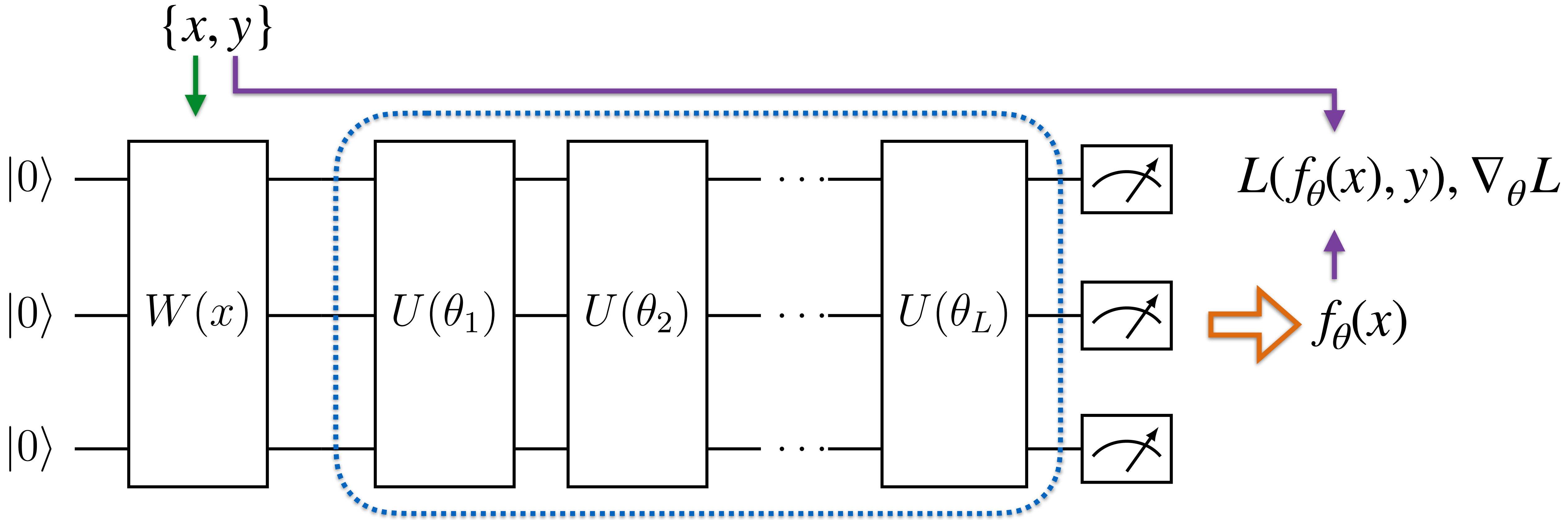
Parametric gate “layers”

Application: Quantum “neural network”



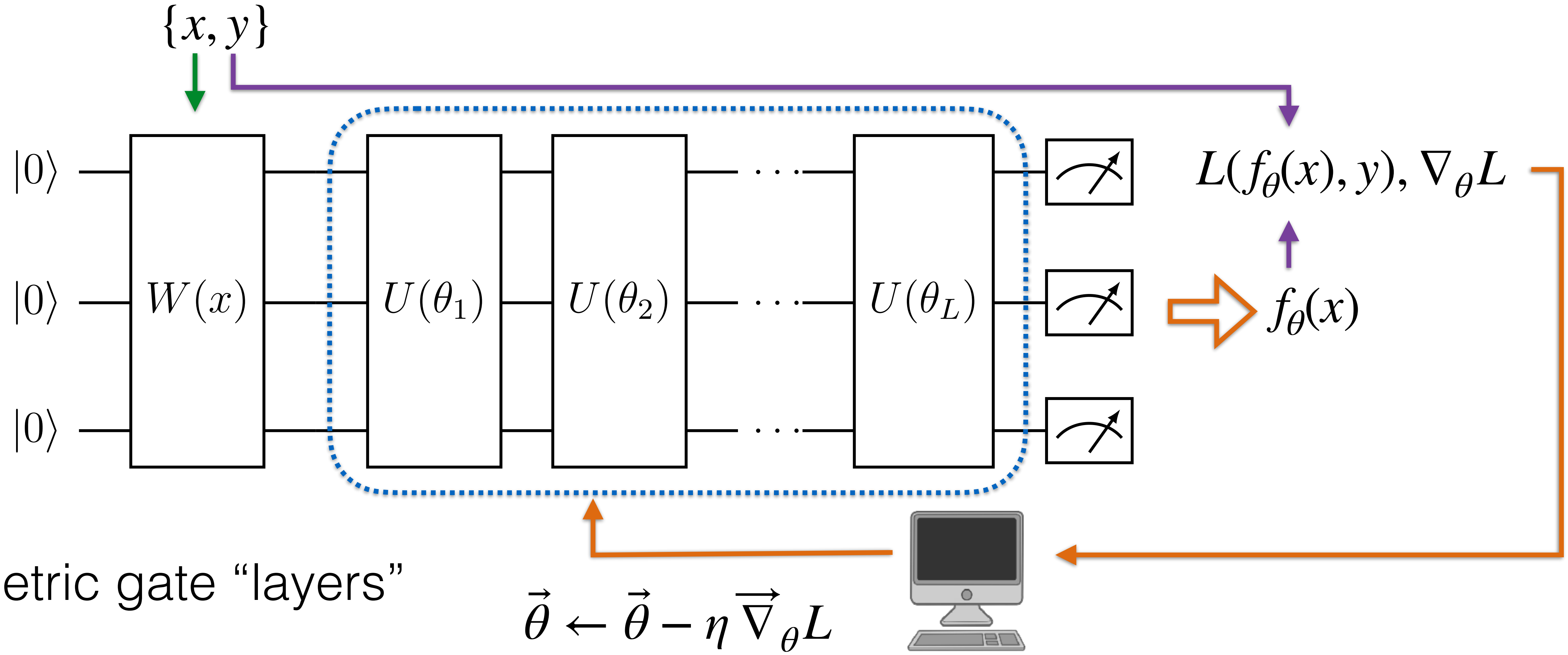
Parametric gate “layers”

Application: Quantum “neural network”



Parametric gate “layers”

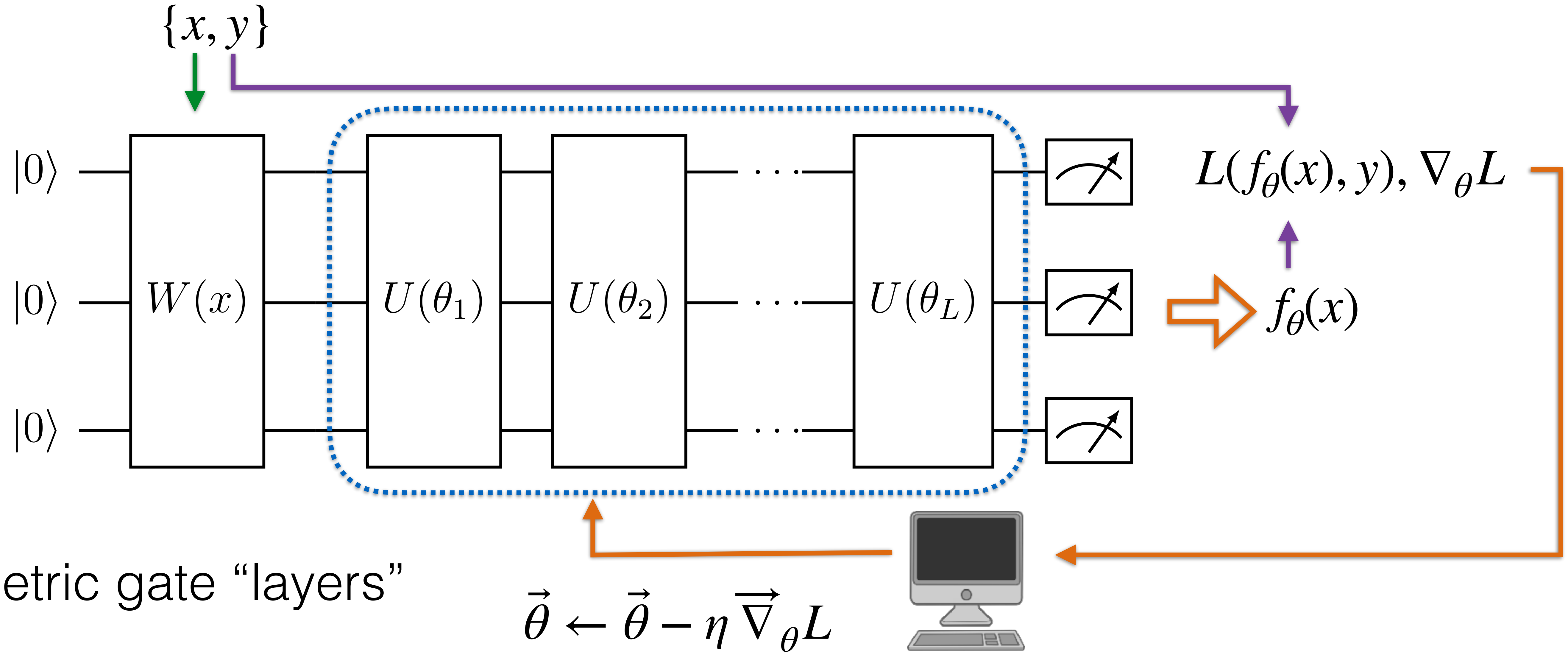
Application: Quantum “neural network”



Parametric gate “layers”

$$\vec{\theta} \leftarrow \vec{\theta} - \eta \overrightarrow{\nabla}_\theta L$$

Application: Quantum “neural network”



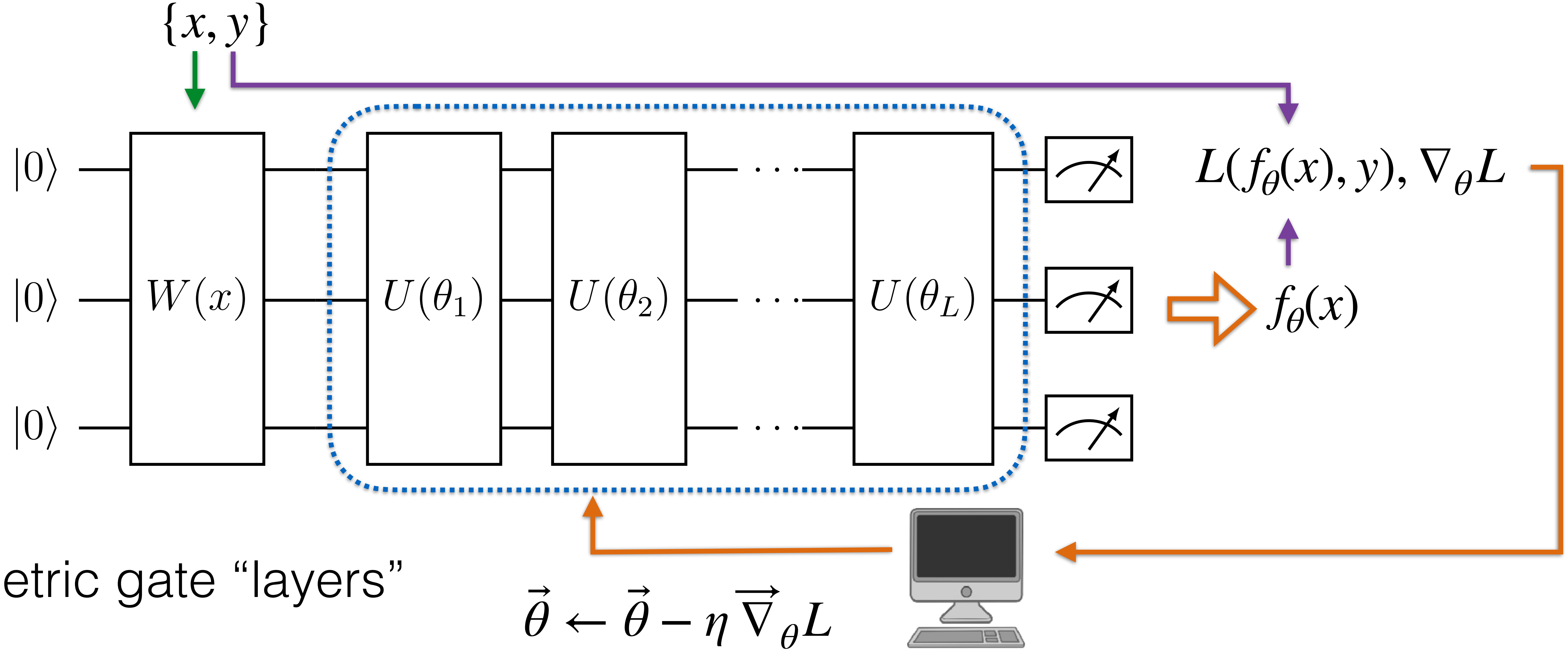
Parametric gate “layers”

$$\vec{\theta} \leftarrow \vec{\theta} - \eta \vec{\nabla}_\theta L$$

→ Quantum-classical hybrid computation

→ Quantum part is short and simple; near-term feasible

Application: Quantum “neural network”



Parametric gate “layers”

$$\vec{\theta} \leftarrow \vec{\theta} - \eta \vec{\nabla}_\theta L$$

Quantum advantage?
Very active research ongoing

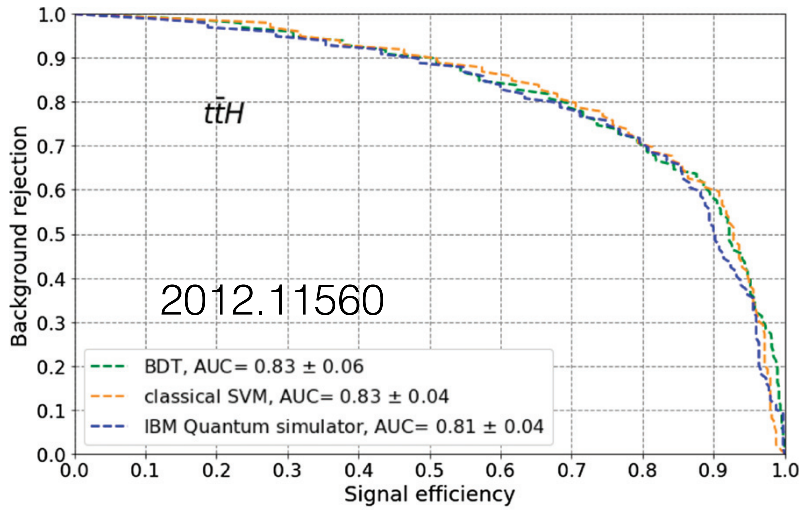
- → Quantum-classical hybrid computation
- → Quantum part is short and simple; near-term feasible

Early HEP adopters

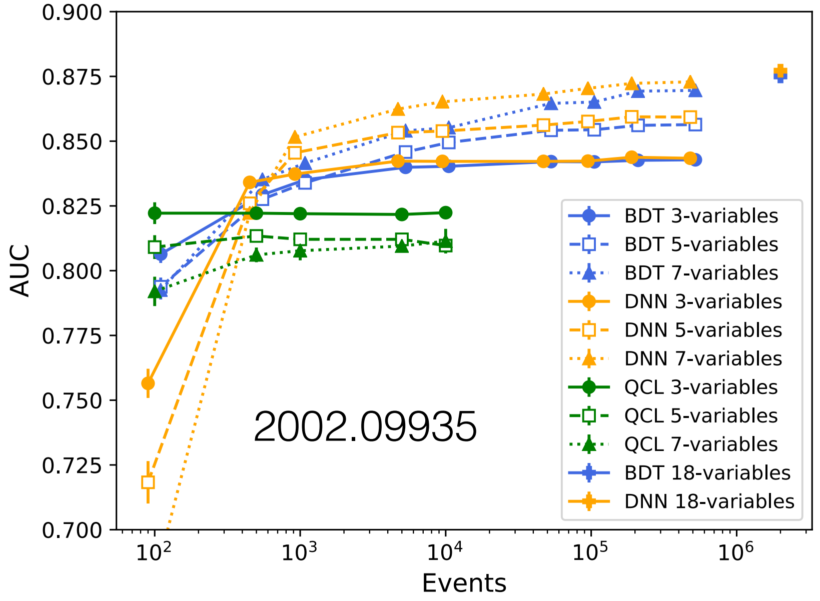
- Wu et al. (JPhys G 2021)
  - Classification of LHC ttH&Hμμ events vs background
- Terashi et al. (CSBS 2021)
  - Classification of LHC SUSY events
- Blance and Spannowsky (JHEP 2020)
  - Classification of Z'→tt against continuum bkg.
- Gianelle et al. (JHEP 2022)
  - b-jet charge identification

Quantum GAN with parametric circuits:

- Chang et al. (ACAT 2021)
  - Calorimeter simulation
- Bravo-Prieto et al. (Quantum 2022)
  - Event generation

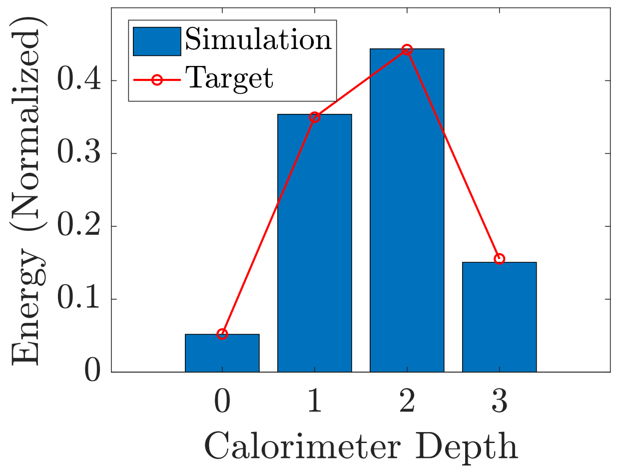


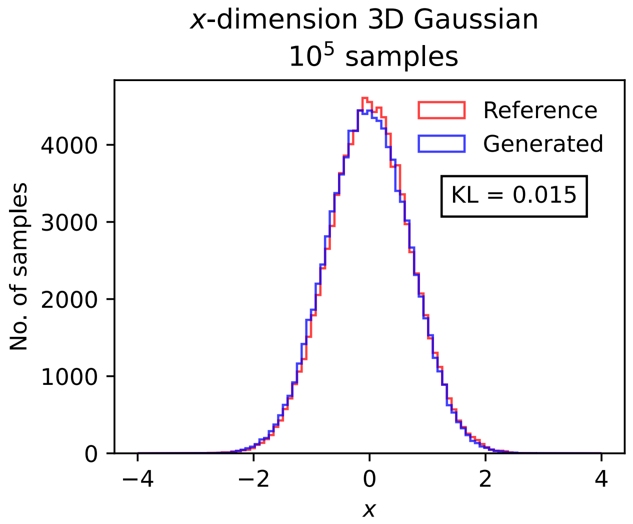
2012.11560

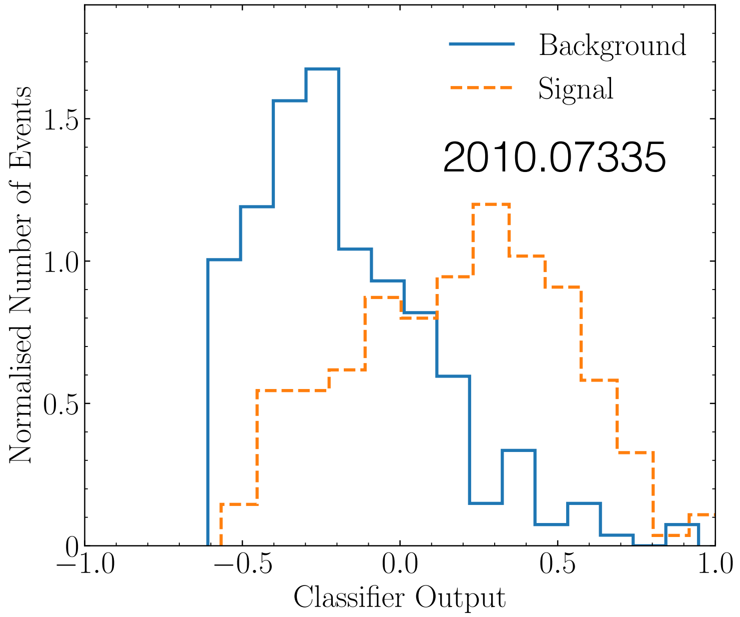


2002.09935

2103.15470







2010.07335

Summary

- Computing and software evolves fast
  → Investment in cutting-edge technologies pays off
- Advanced computing in 2023: heterogeneous, accelerated
  - Driven by need: cannot handle HL-LHC by stacking commodity CPU
  - Also by benefits: novel techniques bring new opportunities
- Reviewed some technologies; left out a lot more
- Noisy quantum computers are available for use
- Exciting outlook for error-corrected quantum computers
  - Also exciting to identify the use cases of noisy devices