Scalable volumetric benchmarks based on Clifford and free-fermion operations

Orsolya Kálmán

Quantum Computing and Information Research Group, HUN-REN Wigner RCP

November 27, 2025

AIME 2025, Budapest

Research directions of the group

- Characterization of errors in quantum computers, benchmarking performance
- Quantum machine learning problems (including classical training, but quantum sampling)
- Photonic quantum computing (including ML, and QML)
- Tensor network simulations
- Distillation schemes for quantum networks or distributed quantum computing

Involvement in projects

OpenSuperQPlus

QUANTERA

SPINUS

eureka

Benchmarking quantum computers

- What is a **Quantum Computer Benchmark**?
  - Quantum computer **benchmarks** are structured methods specifically designed to evaluate quantum system performance.
  - A benchmark defines a set of computational tasks for the quantum processor to execute and prescribes how to analyse the resulting data to compute a benchmark score.

Types of benchmarks:

- **Component-level benchmarks**
  Measure the quality of individual components: gate fidelities, T1/T2 times, measurement errors.

- **Device-level benchmarks**
  Measure the entire processor's ability to run circuits: volumetric benchmarks, layer fidelities.

- **Algorithmic benchmarks**
  Compare the performance of entire families of protocols executing algorithms.

Quantum Benchmarks for different QC eras

Era of Quantum Computing	NISQ	late-NISQ	Early Fault-Tolerant	Full Fault-Tolerant
Typical gate errors	10^{-1} - 10^{-3} (physical 2Q-gate)	10^{-3}-10^{-5} (physical 2Q-gate)	10^{-5} - 10^{-8} (logical gate)	10^{-8} - (logical gate)
Feasible number of gate operations	10 - 1 000	1 000 - 100 000	1 000 - 1M	>1M
Component-level benchmarks	Very relevant	Very relevant	Very Relevant	Relevant
Device-level benchmarks	Relevant	Very Relevant	Relevant	Relevant
Algorithmic benchmarks	Less relevant	Less relevant	Relevant	Very Relevant

Benchmarking quantum computers (device level)

Near-term benchmarking approach: Instead of full-fledged algorithmic benchmarks, it is more natural to consider benchmarks testing algorithmic primitives, i.e., protocols that appear as subtasks within algorithms.

- Rather than complete algorithmic benchmarks, we focus on benchmarks **evaluating algorithmic primitives**—smaller-scale protocols and subroutines crucial to quantum algorithms.
- This approach offers clear insight into **how close we are to achieving practical quantum utility** without requiring the execution of a full-scale quantum algorithm.
- It enables **tracking** and guiding **the evolution of quantum computational platforms** toward achieving genuine utility.

Criteria for device-level (volumetric) benchmarks

Benchmarks should satisfy (ideally):

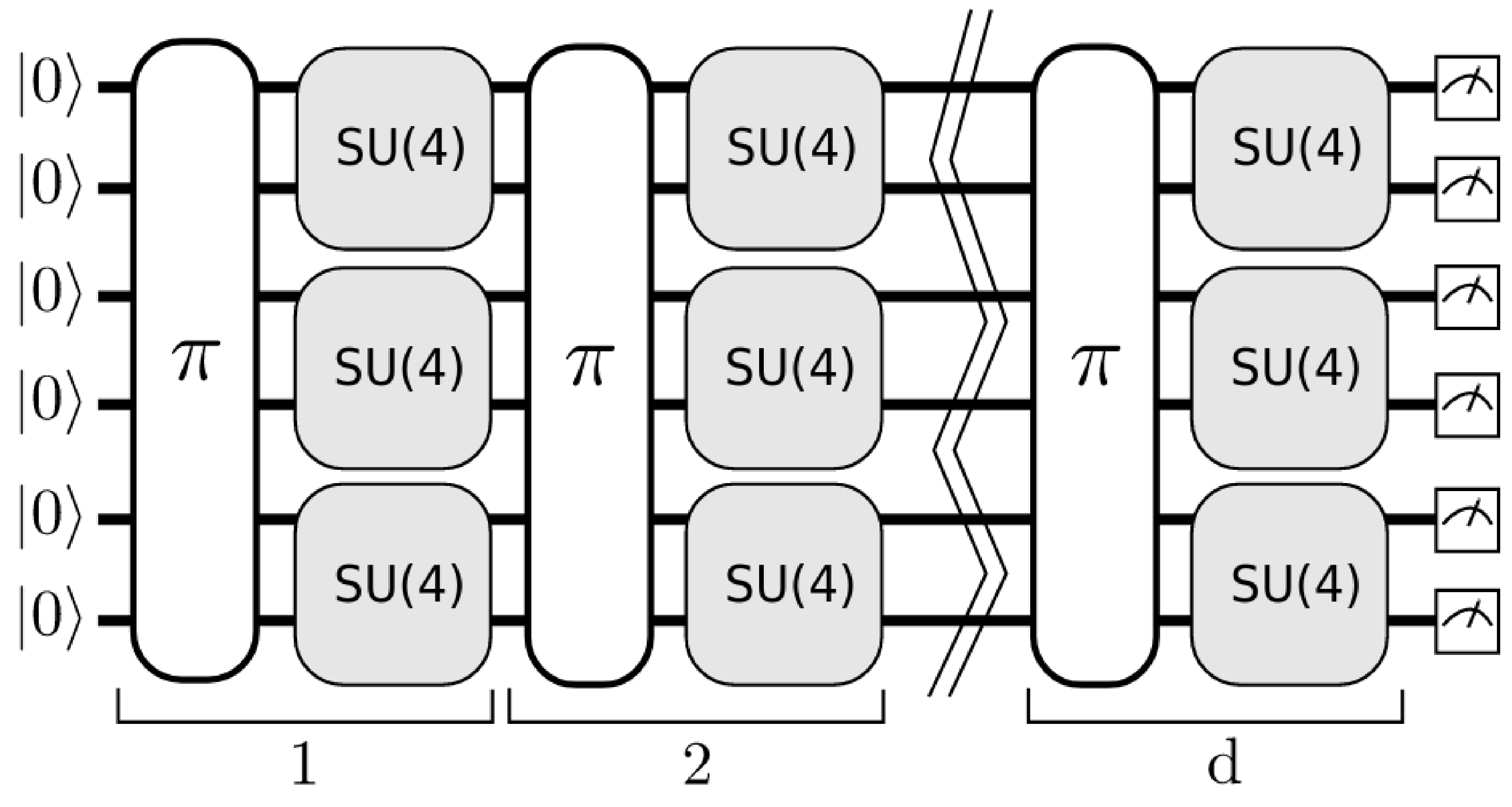
- Well-defined: clear set of rules, reproducible
- Platform independent, hardware-agnostic
  - no native gates, no fixed connectivity
  - preferably the unitary is given as a suitably defined global operation with no particular reference to compilation
- Scalable (wrt the circuit sizes relevant for the era)

Additional considerations:

- Application oriented (relevant for the era)
- Fit to the present devices (e.g., number of shots, software tools)
- Well known, often used by companies, academics
- Randomness involved
- Interesting science involved

Quantum Volume

- Quantum Volume uses **random square circuits with all-to-all connectivity**
  - d sequential layers acting on d qubits
  - a random permutation of these qubits
  - Haar-random SU(4) operations performed on neighboring pairs



A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, J. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A 100, 032328 (2019).

- **Definition of Quantum Volume**

$$\log_2 V_Q = \underset{m}{\operatorname{argmax}} \min(m, d(m))$$

- *m*: number of qubits
- *d(m)*: the number of qubits in the largest square circuits that can reliably sample heavy outputs with p > 2/3

Advantages

- It was well tailored for the NISQ era: fit to the present devices, software tools
- Well known, often used by companies, academics
- Interesting theory/physics involved

Disadvantages

- Not scalable for the late-NISQ (and the Fault Tolerant) era
- The value is not natural (an artificial exponential involved)
- Partially platform biased
- Not application oriented

Our volumetric framework

- Computational Task:
  - Algorithmic primitives
  - Quantum operations that remain **classically simulable**.
  - Trade-off between scalability and quantum utility demonstration.
  - Abstract definition for flexible compilation and optimization, ensuring platform independence.
- Measurement Procedure:
  - Measure specifically chosen expectation values to verify correct quantum state preparation in noise-free conditions.
  - Success criteria
- Benchmark Score:
  - Largest number of qubits on which the quantum processor consistently meets success criteria.

Clifford Volume Benchmark

Initialization and Circuit Preparation:

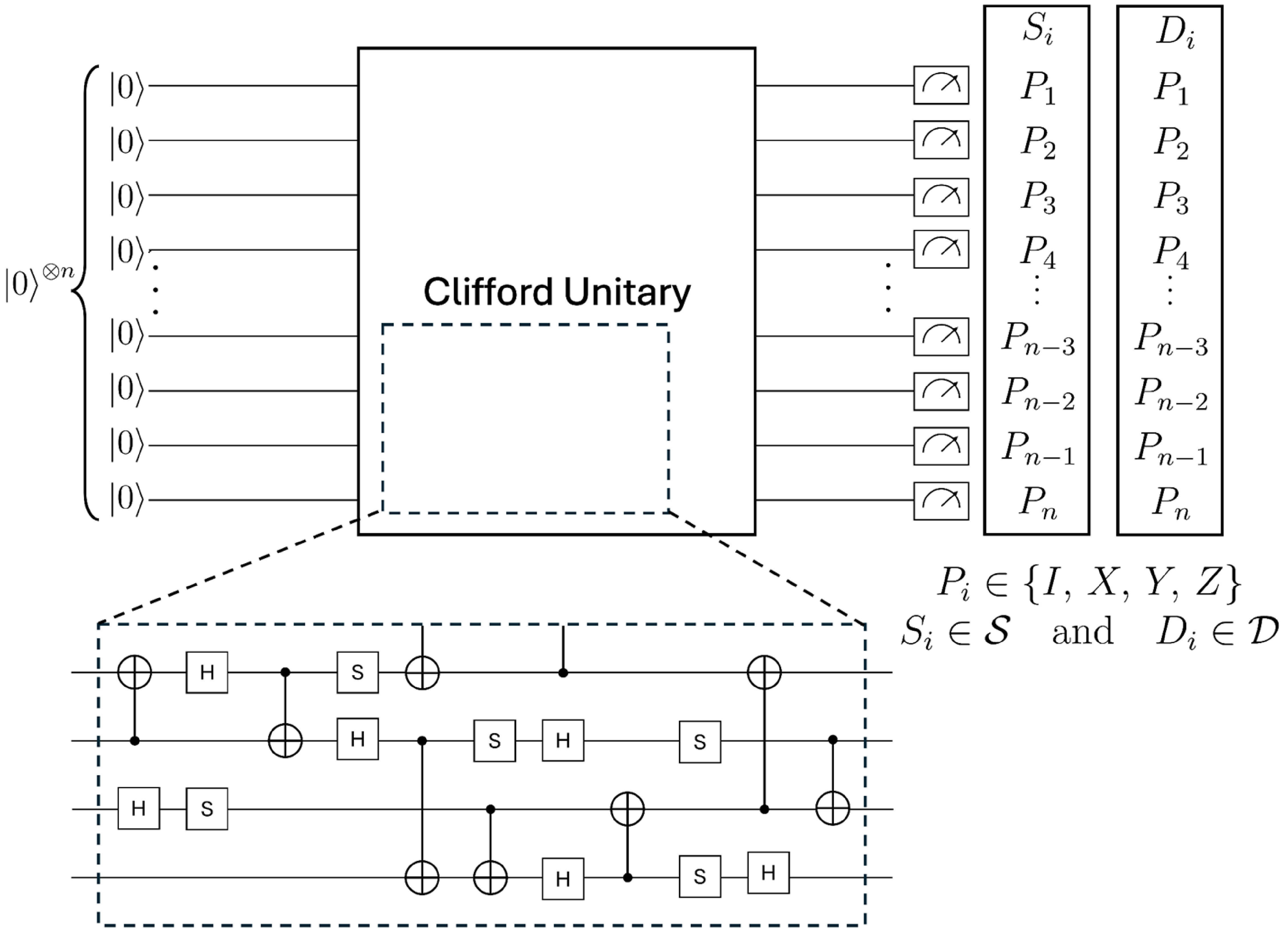
The following steps are carried out by the classical part of the quantum computational stack.

1. Randomly select **M=10** unitaries from the ***n*-qubit Clifford group.**

2. For each selected Clifford operations, determine ***n* generators of the stabilizer group**, additionally, randomly select n distinct ***n*-qubit Pauli operators that lie outside the stabilizer group**.

3. For ***n* > 10**, we restrict the set of generators by randomly selecting **$10 + \lfloor n / 10 \rfloor$** stabilizer operators and an equal number of non-stabilizer operators from the full set.

4. **Construct and compile the circuit implementation** of each n-qubit Clifford, optimised for the benchmarked platform.



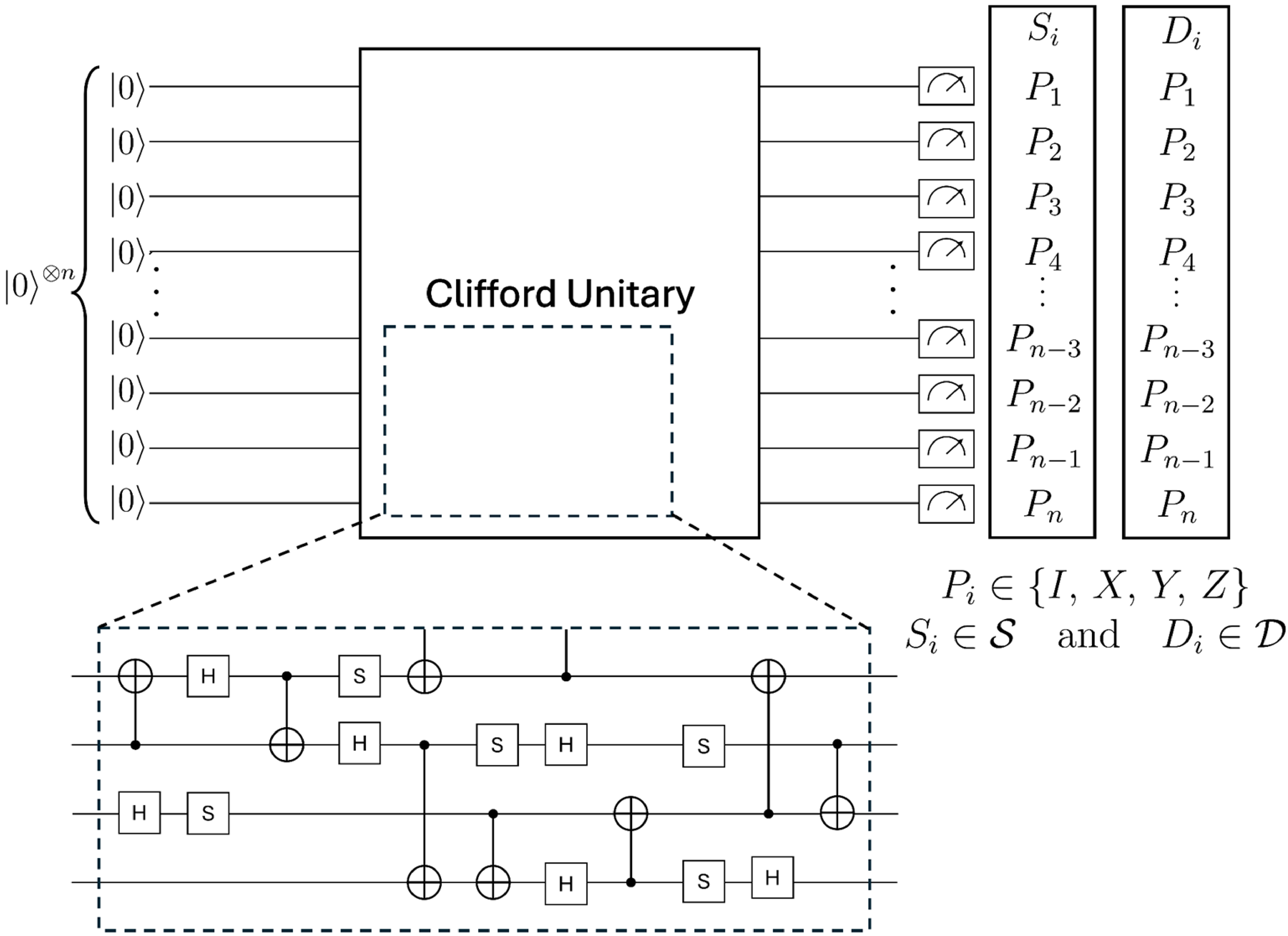
Clifford Volume Benchmark

Circuit Execution:

This step must be carried out on a quantum device.

For each Clifford unitary operation

1. Initialize all qubits in the $|0\rangle$ state.
2. Execute the circuit implementation 2^{12} times for each selected **stabilizer** and **non-stabilizer operators.**
3. Measure the **expectation value** of the selected **stabilizer** and **non-stabilizer operators.**



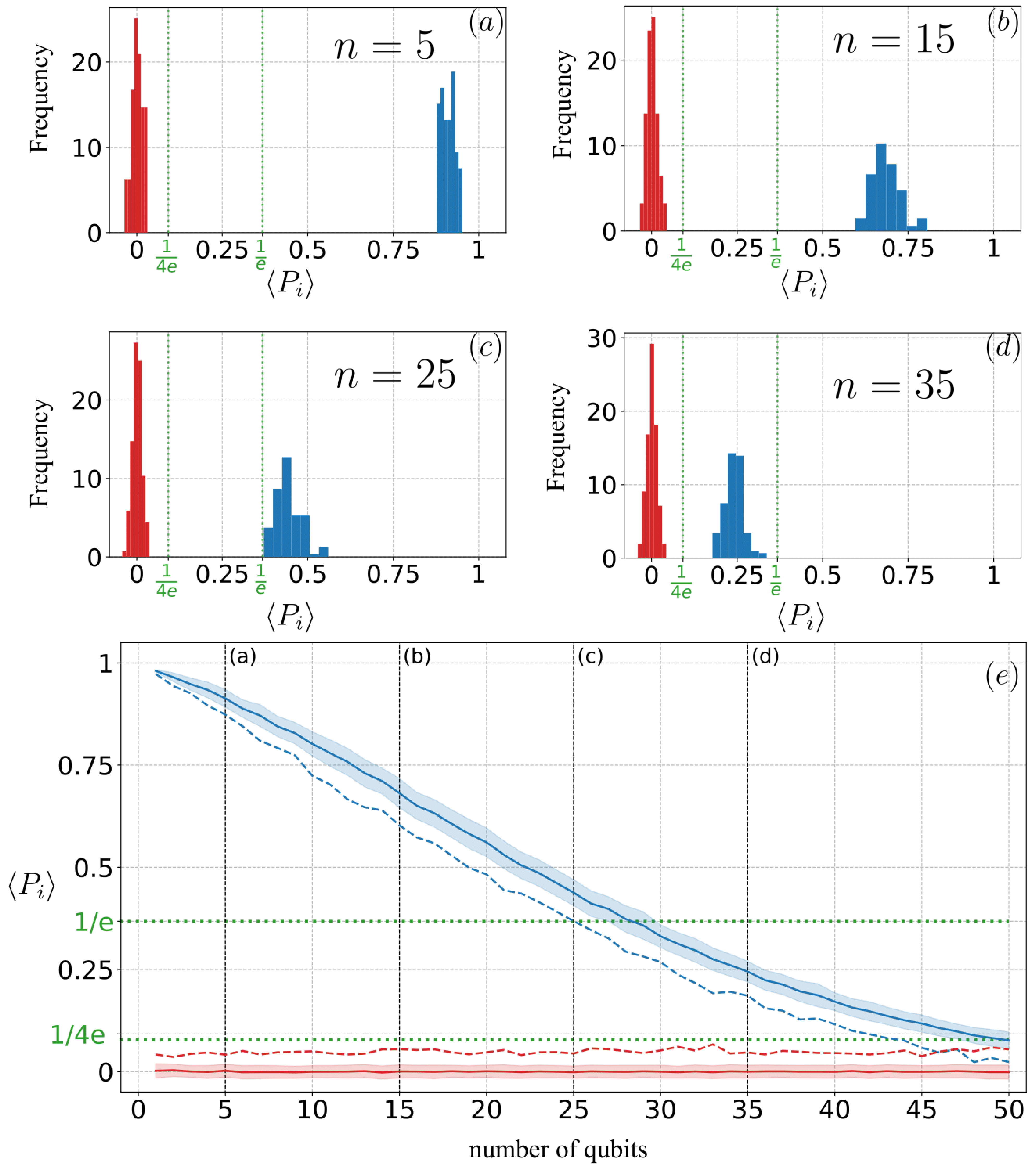
Clifford Volume Benchmark

Performance Evaluation: We consider the ***n***th step of the benchmark successful if the following conditions are fulfilled simultaneously:

$$\min_{m,i} \langle S_i^m \rangle \geq \frac{1}{e}$$ for every chosen Clifford unitary and corresponding stabilizer generator

$$\max_{m,i} |\langle D_i^m \rangle| \leq \frac{1}{4e}$$ for every chosen Clifford unitary and corresponding non-stabilizer operator

Benchmark score: The largest ***n*** value for which the benchmark protocol succeeds serves as a metric to characterize the quantum computer.



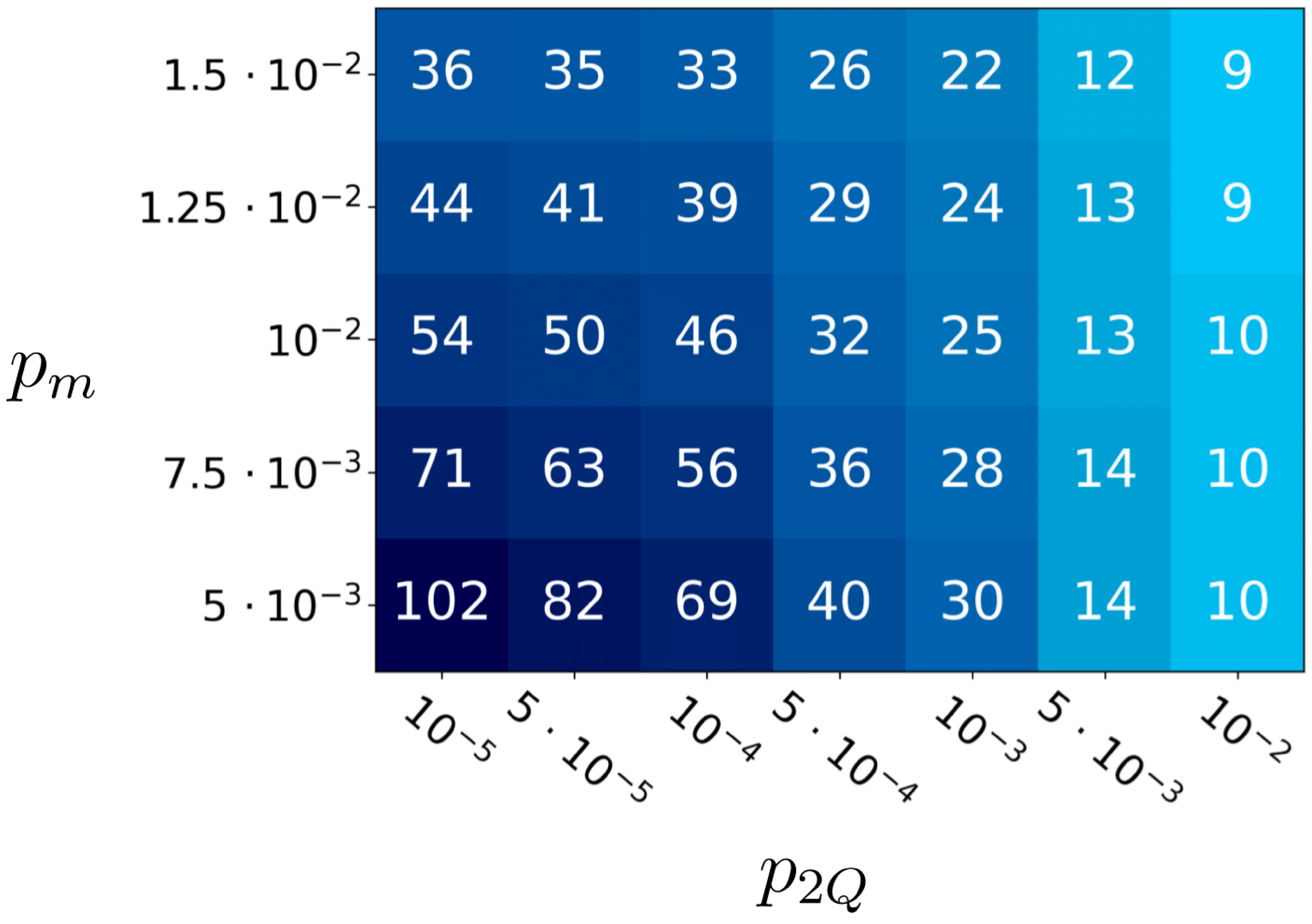
$$p_m = 10^{-2},\ p_{2Q} = 10^{-3}$$

Clifford Volume Benchmark

Performance Evaluation: We consider the ***n***th step of the benchmark successful if the following conditions are fulfilled simultaneously:

$$\min_{m,i} \langle S_i^m \rangle \geq \frac{1}{e}$$ for every chosen Clifford unitary and corresponding stabilizer generator

$$\max_{m,i} |\langle D_i^m \rangle| \leq \frac{1}{4e}$$ for every chosen Clifford unitary and corresponding non-stabilizer operator



Benchmark score: The largest ***n*** value for which the benchmark protocol succeeds serves as a metric to characterize the quantum computer.

Clifford Volume Benchmark

Advantages:

- **Platform-independence:** The unitaries are selected from a group of global operations, we avoid prescribing specific compilation implementations.
- **Application-oriented**: Clifford circuits form the backbone of many fault-tolerant protocols, Shadow tomography, energy estimations,
- **Scalability**: Classically simulable operations, validates the benchmark results using stabilizer fidelity witnesses, ensuring the benchmark remains scalable.

Disadvantage:

- Clifford operations alone cannot achieve quantum advantage, but ...

Free-fermion Volume Benchmark

Free fermionic systems are defined by their quadratic Hamiltonians of the form

$$H = \frac{i}{4}\sum_{j,k} A_{j,k} m_j m_k$$

where A is a real antisymmetric matrix, and m_j are the Majorana operators.

The corresponding unitary operation is given by

$$F(t) = \exp\left(\frac{1}{4}\sum_{i,j=1}^{2d} [\log(O(t))]_{ij} m_i m_j\right)$$

where

$$O(t) = e^{tA} \in SO(2L)$$

Considering the time evolution governed by the free-fermion operator, we can verify the orthogonality relation for the matrix O by measuring the expected values of the corresponding Majorana mode operators, given by

$$\langle m_j \rangle_{F\rho_i F^\dagger} = \mathrm{Tr}\left(F\rho_i F^\dagger\, m_j\right) = \sum_k O_{jk} \langle m_k \rangle_{\rho_i} = \sum_k O_{jk}\delta_{ik} = O_{ji}$$

This allows us to test the orthogonality relation by measuring the expected values of the Majorana operators:

$$\sum_k O_{kj} \langle m_k \rangle_{F\rho_i F^\dagger} = \sum_k O_{ki} O_{kj} = \delta_{ij}\ .$$

Free-fermion Volume Benchmark

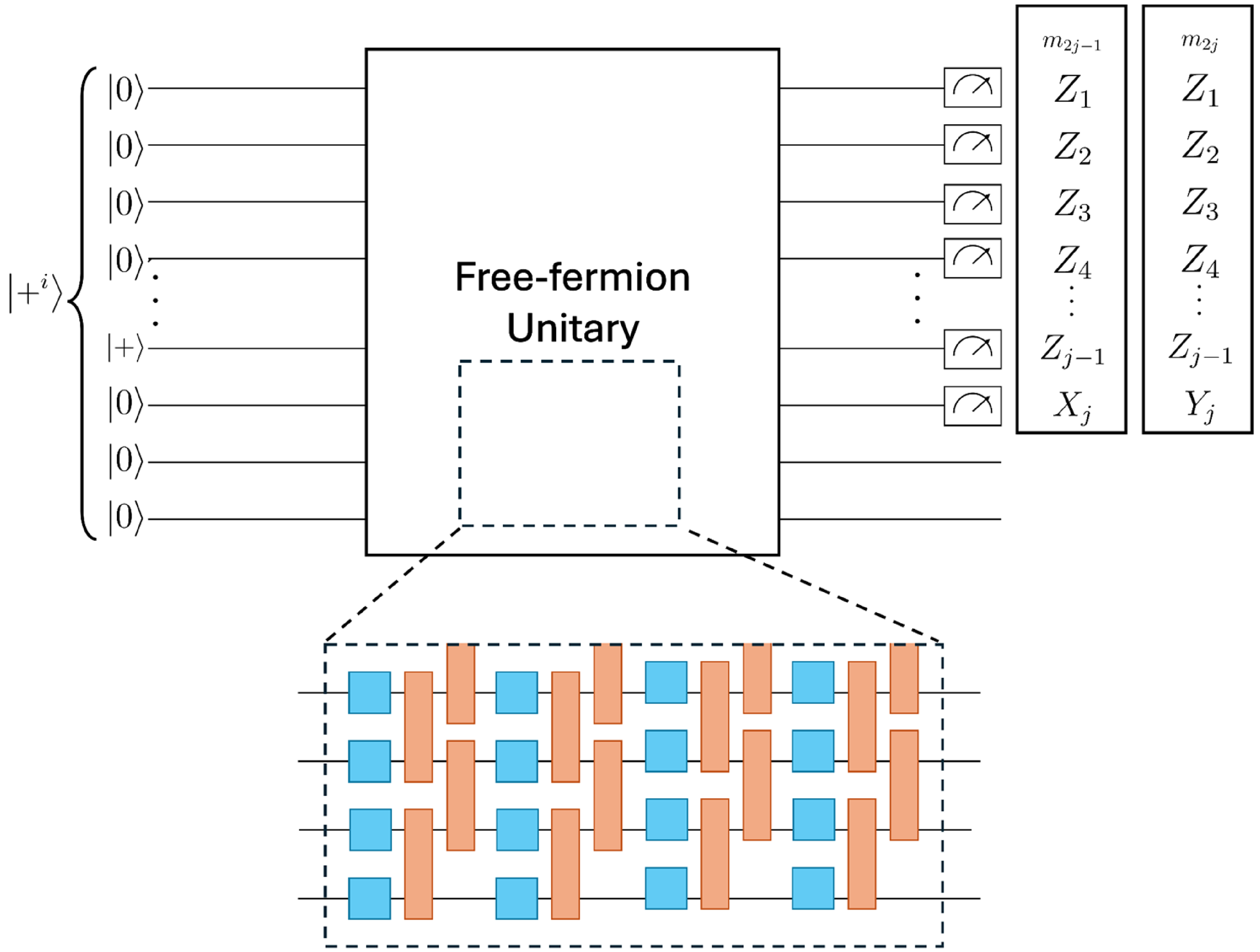
Initialization and Circuit Preparation:

The following steps are carried out by the classical part of the quantum computational stack.

1. Randomly select **M=10** matrices from the **SO(2*n*)** group corresponding to ***n*-qubit free-fermion unitaries (matchgates).**

2. If ***n* > 10**, we select the **20 + ⌊2*n* / 10⌋** largest expectation values determined by the chosen **O matrices** of the SO(2*n*) group, these determine the **Majorana operators** to be measured.

3. **Construct and compile** a circuit implementation of each free-fermion unitary.



Free-fermion Volume Benchmark

Circuit Execution:

The following steps must be carried out on a quantum device.

For each Free-fermion unitary

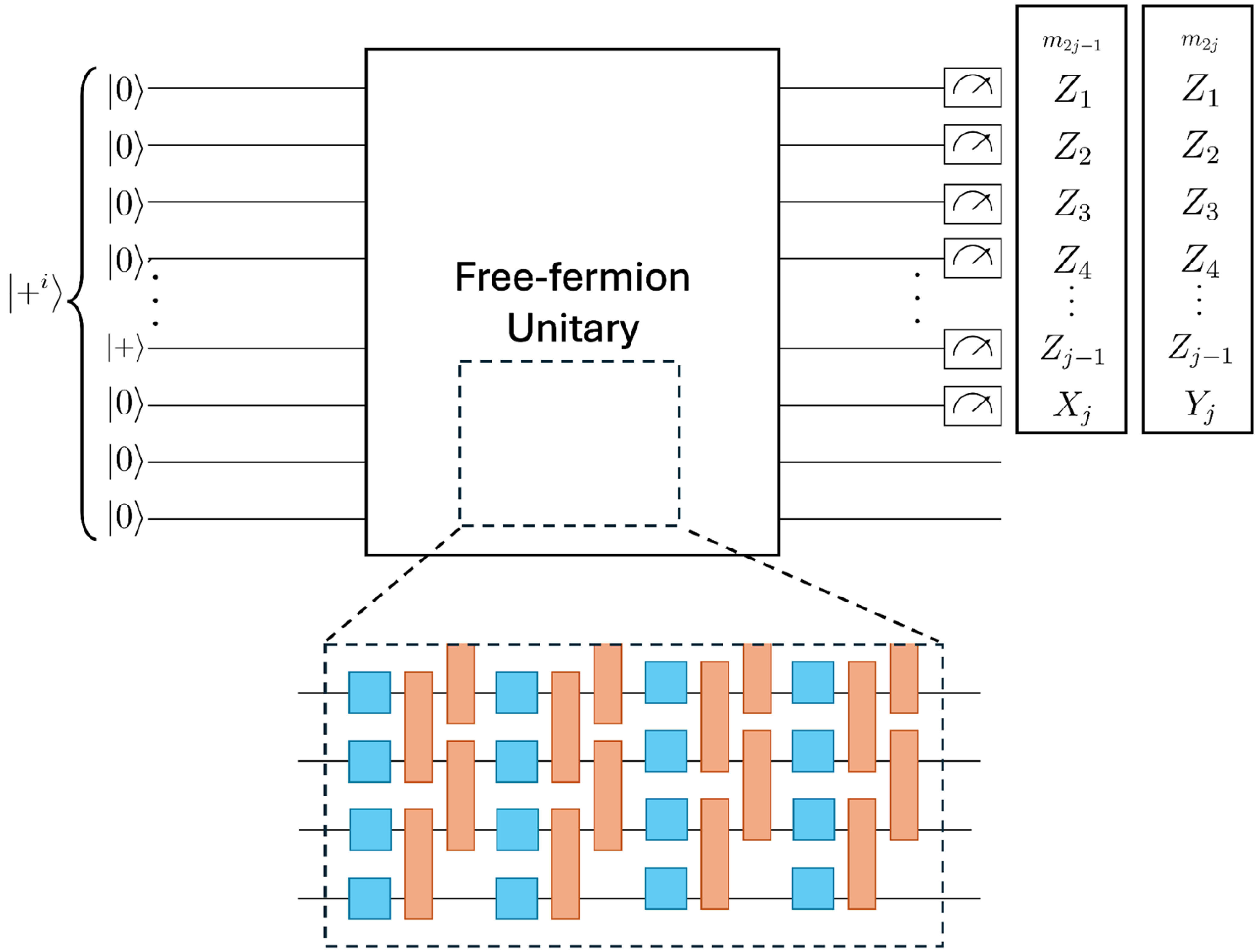
1. Initialize the system in a randomly chosen uniform superposition of fermionic states:

$$\rho_i = |+^i\rangle\langle +^i| \quad \text{where} \quad |+^i\rangle = |0\rangle^{\otimes i-1} \otimes |+\rangle \otimes |0\rangle^{\otimes n-i}$$

2. Execute the circuit implementation.

3. Measure the expectation value of each chosen Majorana operators.

$$m_{2p-1} = Z_1 Z_2 \cdots Z_{p-1} X_p, \quad m_{2p} = Z_1 Z_2 \cdots Z_{p-1} Y_p$$



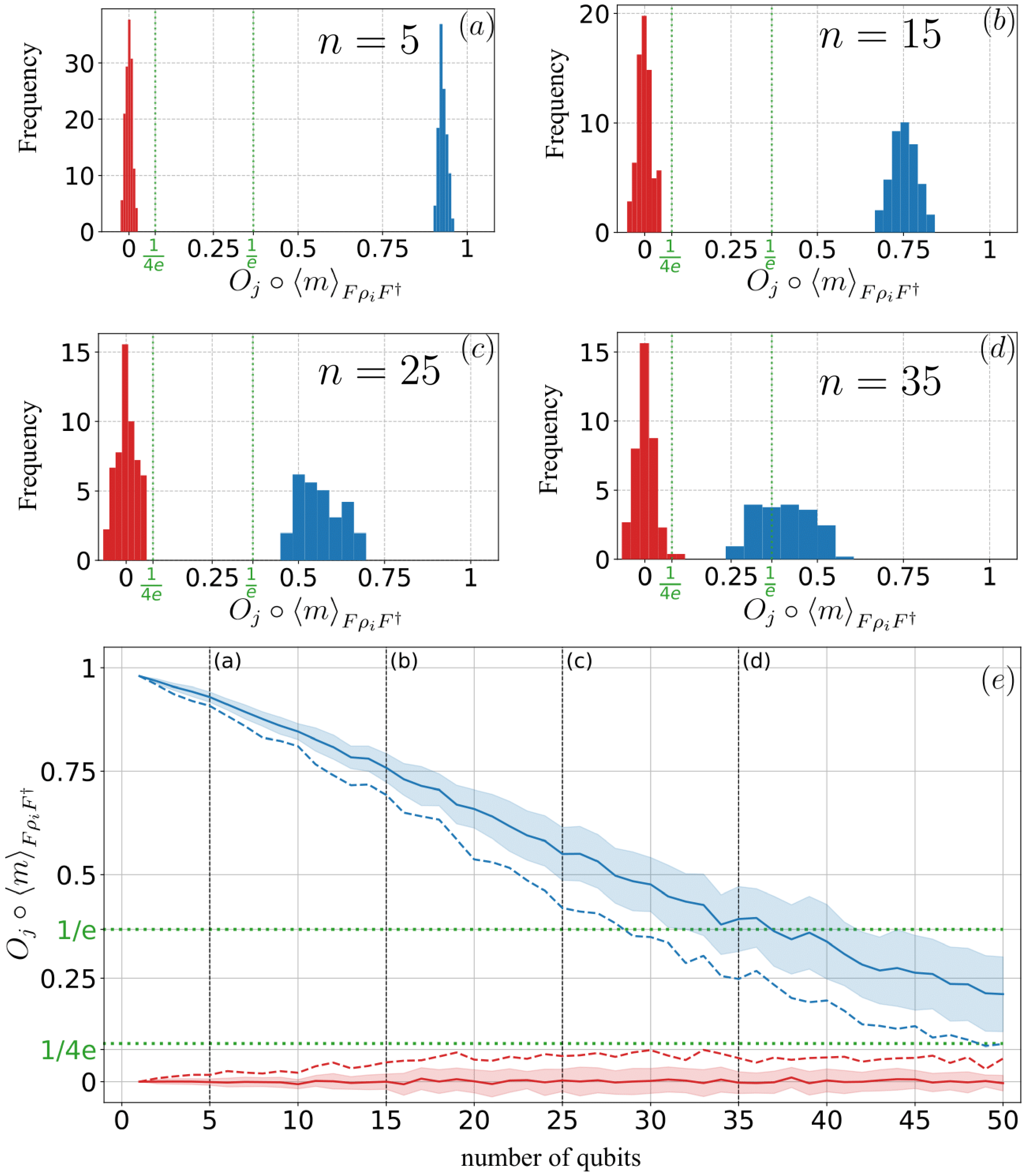
Free-fermion Volume Benchmark

Performance Evaluation: We consider the ***n***th step of the benchmark successful if the following conditions are fulfilled simultaneously:

$$\min_q \left(\sum_k O_{ki}^{(q)} \left\langle m_k^{(q)} \right\rangle \right) \geq \frac{1}{e}$$

$$\max_q \left(\sum_k O_{kj}^{(q)} \left\langle m_k^{(q)} \right\rangle \right) \leq \frac{1}{4e} \qquad \text{for } j \neq i$$

Benchmark score: The largest ***n*** value for which the benchmark protocol succeeds serves as a metric to characterize the quantum computer.



$p_m = 10^{-2},\ p_{2Q} = 10^{-3}$

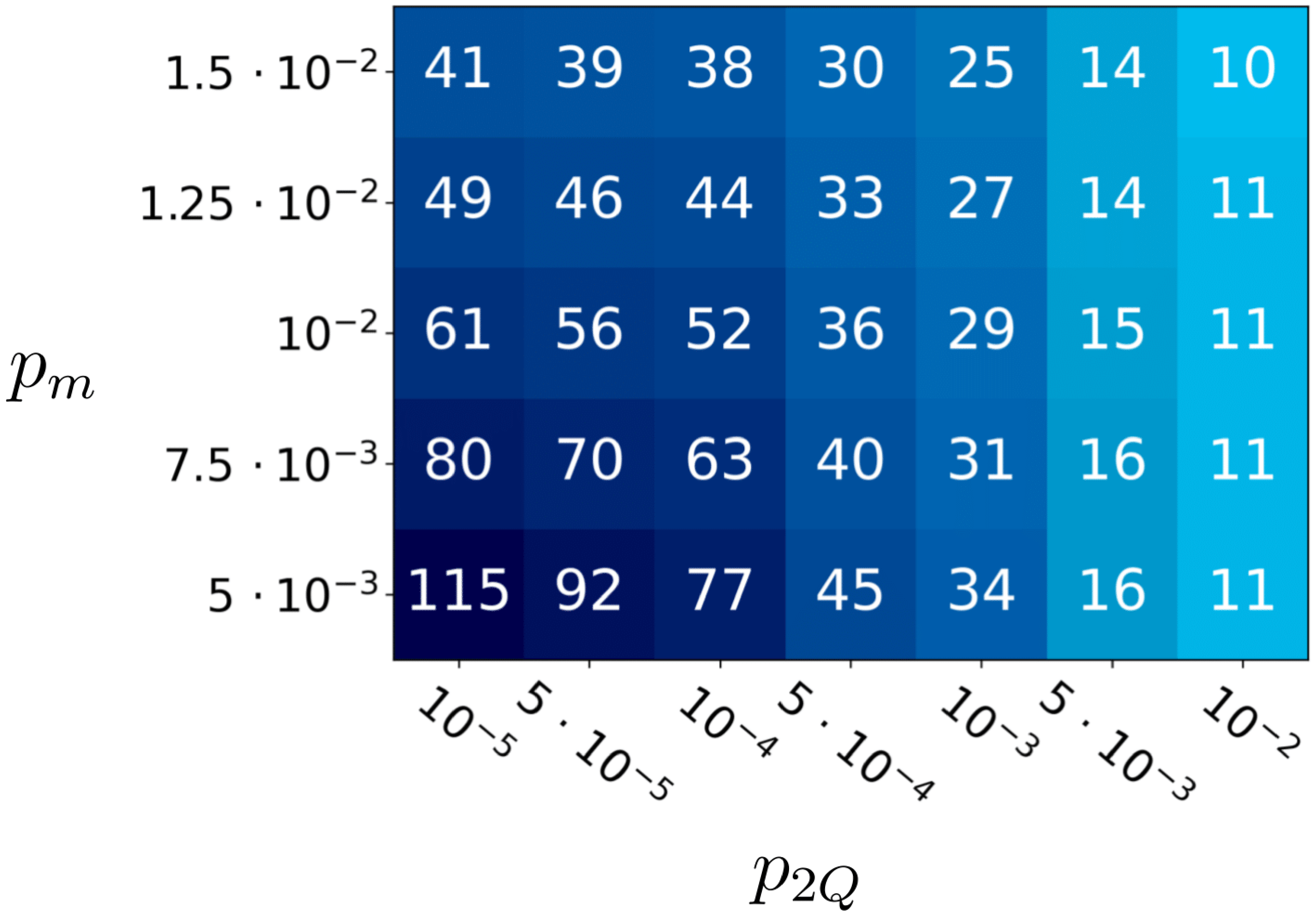
Free-fermion Volume Benchmark

Performance Evaluation: We consider the ***n***th step of the benchmark successful if the following conditions are fulfilled simultaneously:

$$\min_q \left(O^{(q)} \left\langle m_k^{(q)} \right\rangle \right) \geq \frac{1}{e} \qquad \text{for k = i}$$

$$\max_q \left(O^{(q)} \left\langle m_k^{(q)} \right\rangle \right) \leq \frac{1}{4e} \qquad \text{for k} \neq \text{i}$$

Benchmark score: The largest ***n*** value for which the benchmark protocol succeeds serves as a metric to characterize the quantum computer.



Free-fermion Volume Benchmark

Advantages:

- **Platform-independence:** The unitaries are selected from a group of global operations, we avoid prescribing specific compilation implementations.
- **Application-oriented**: energy estimations, fermion-to-qubit mapping changes, simulation of free-fermion Hamiltonians, quantum chemistry, quantum many-body models
- **Scalability**: Classically simulable operations, validates the benchmark results using fidelity witnesses, ensuring the benchmark remains scalable.
- **Together with Clifford** unitaries they form a **universal gate-set**, thus when applied jointly they are hard to simulate classically. Example: **Fermion Sampling**, which can be validated by measuring the 4-point correlators while sampling from the output state is classically hard.

M. Oszmaniec, N. Dangniam, M. E. S. Morales, Z. Zimborás, Fermion sampling: a robust quantum computational advantage scheme using fermionic linear optics and magic input states, PRX Quantum, 2022.

From the European Quantum Flagship’s KPI Booklet

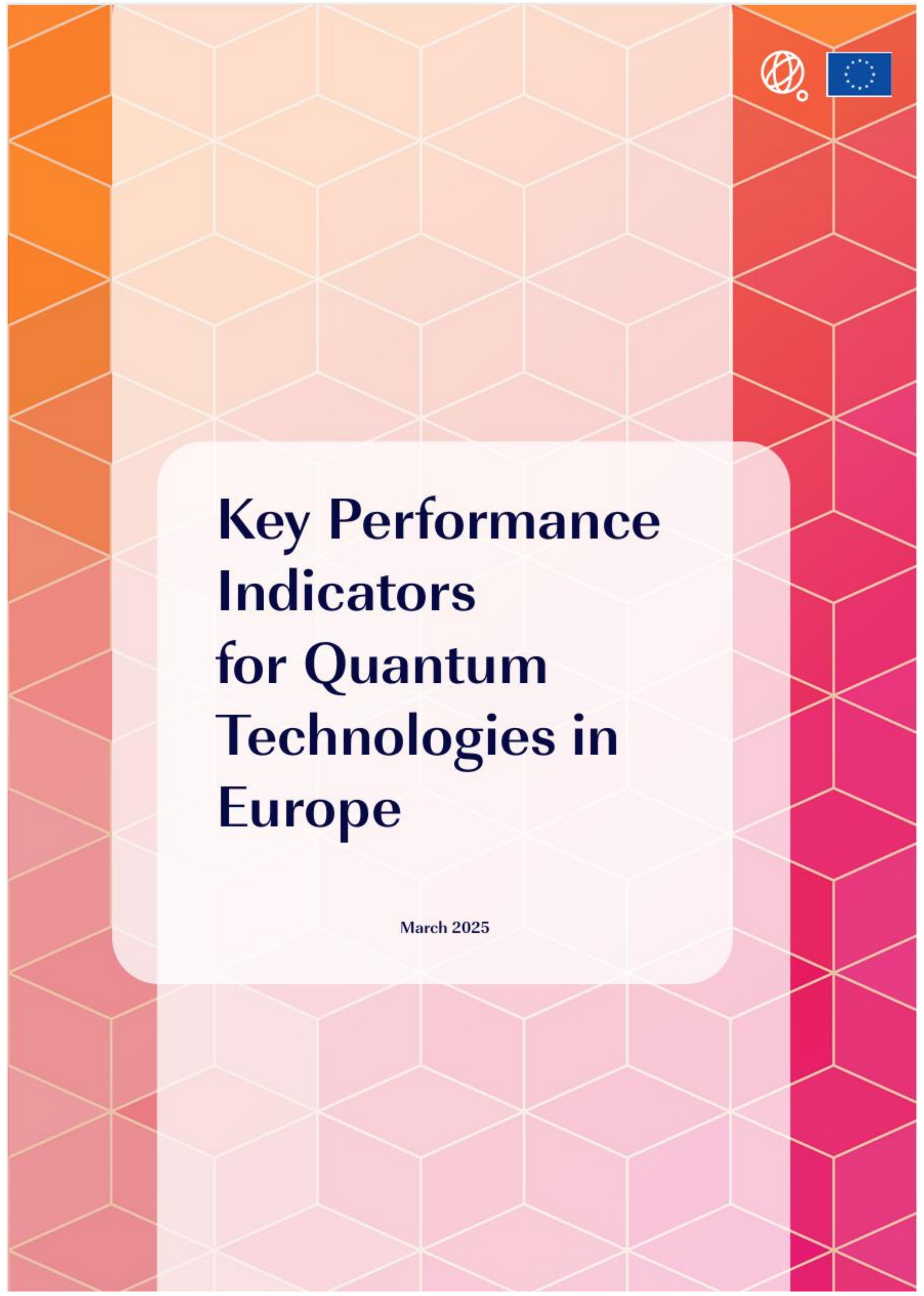
Key Performance Indicators for Quantum Computing

1. **Circuit size:** The largest random N-qubit Clifford operation that can be reliably executed

2. **Multipartite Entanglement:** The largest GHZ state successfully prepared and verified by a quantum computer

3. **Cryptanalysis:** Largest reliable instance of Shor's algorithm.

4. **Error correction:** the ratio of the best achievable Bell state fidelity on physical qubits to the best achievable Bell state fidelity on logical qubits on the same hardware



OpenSuperQPlus

https://ec.europa.eu/eusurvey/runner/Quantum_Flagship_KPI_Survey_2024

MILLENION QUANTUM