QESEM: Reliable, Accurate Error Mitigation for Large-Scale Quantum Circuits

Netanel Lindner
CTO, Qedma Quantum Computing
Prof. of Physics, Technion – Israel Institute of Technology

QEDMA

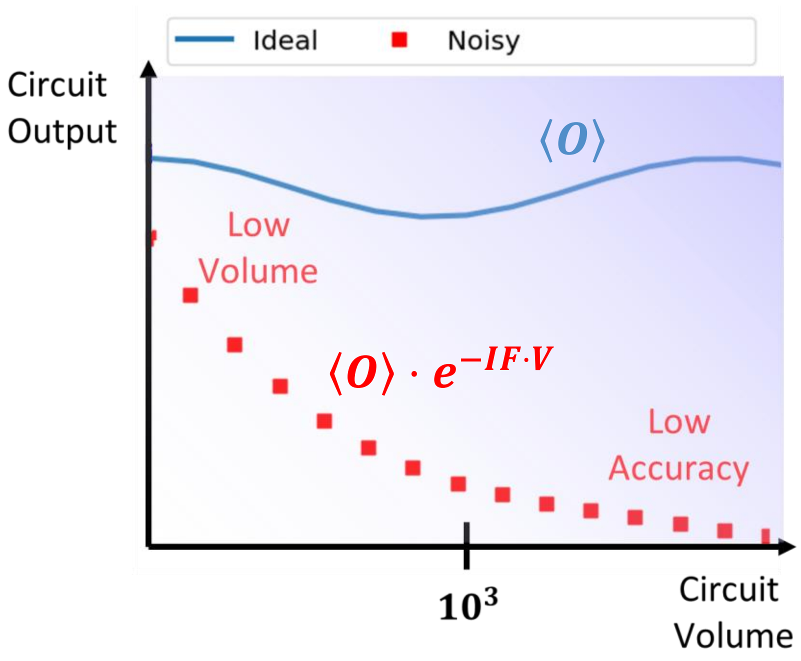
Outline

- **Motivation:** the importance of error mitigation (EM) for quantum computation (based on arXiv: 2503.17243)
  - EM provides dramatic volume boosts over noisy circuit execution
  - EM enables quantum advantage in near-term and error corrected devices
- **QESEM:** Qedma's error mitigation software balancing efficiency and accuracy (based on arXiv: 2508.10997)
  - Workflow overview
  - Largest unbiased EM demo to date on 100 qubits
  - Key innovations behind QESEM
  - QPU-HPC combination using QESEM and classical tensor-network methods
- **Performance projections and outlook**
  - Combining EM with error correction leads to substantial gains
  - Quantum advantage on near term devices

QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Errors – the main obstacle to useful quantum computing



Even a small gate infidelity IF accumulates to a large circuit error

$$(1 - IF)^V \sim e^{-IF \cdot V}$$

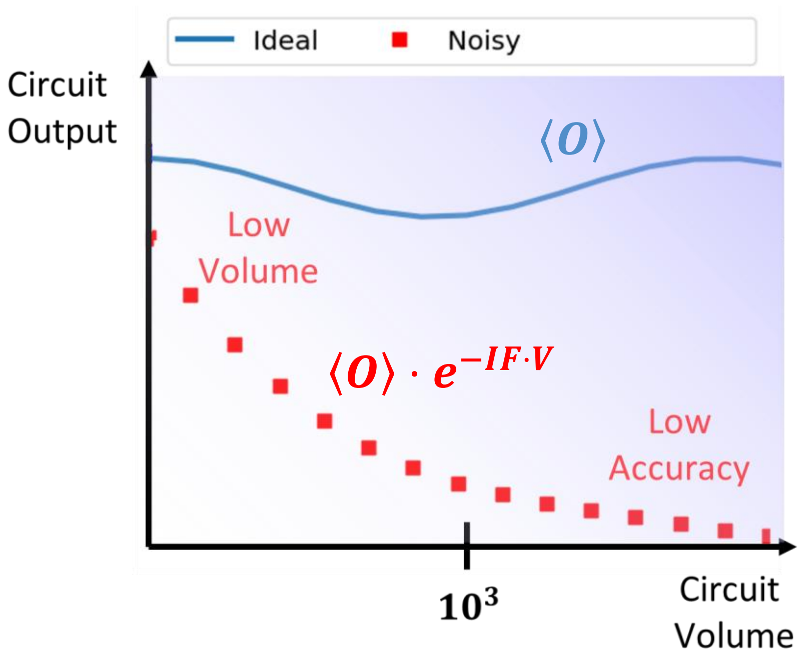
info@qedma.com

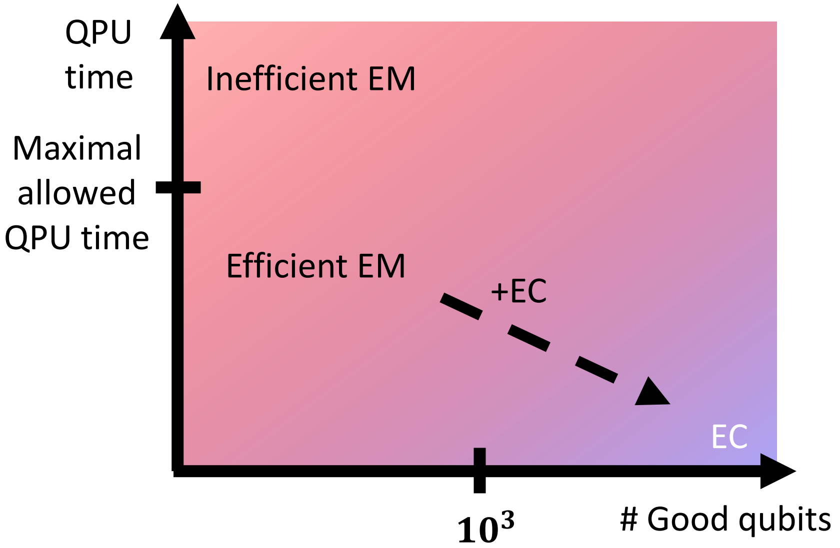
QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Error Correction is not enough, given specs & qubit numbers of near-term QPUs.

Errors – the main obstacle to useful quantum computing





Even a small gate infidelity IF accumulates to a large circuit error

$$(1 - IF)^V \sim e^{-IF \cdot V}$$

Efficient EM is necessary for achieving accurate results both today and as error correction evolves

info@qedma.com

QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Error Correction is not enough, given specs & qubit numbers of near-term QPUs.

Error Mitigation (EM) boosts achievable circuit volumes.

Error mitigation boosts quantum circuit volumes

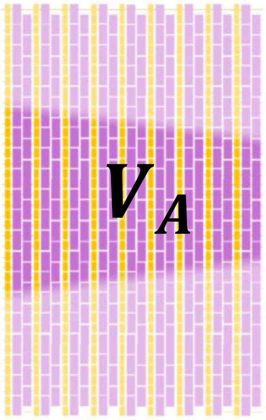
Active volume V_A:

What is it?

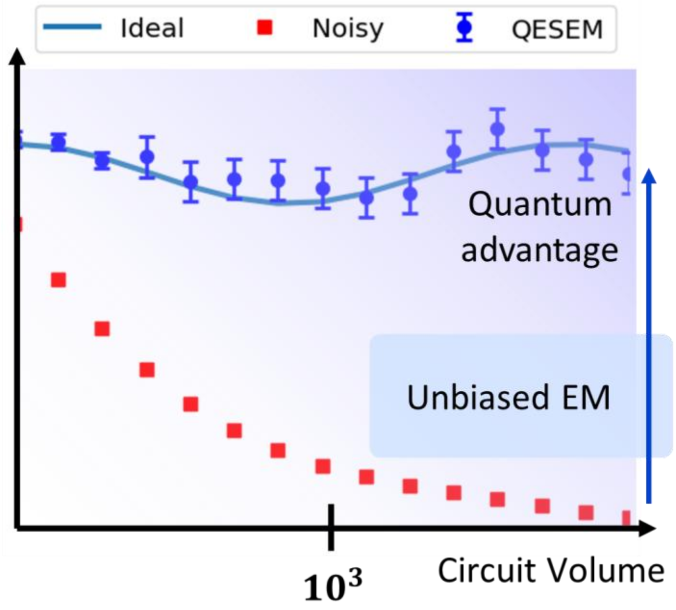
Entangling gates affecting the observable

Why is it important?

Controls the complexity of both error mitigation and classical simulation



$\langle O \rangle$



Errors limit quantum circuit volumes V_A achievable with accuracy ϵ to $V_A \sim \frac{\epsilon}{IF}$

$$\epsilon = \langle O \rangle_{noisy} - \langle O \rangle_{ideal} \sim IF \cdot V_A$$

Unbiased EM boosts circuit volume V_A achievable with accuracy ϵ

$$V_{miti} = \frac{\log(\sqrt{x})}{IF} \Rightarrow \frac{V_{miti}}{V_{raw}} = \frac{\log(\sqrt{x})}{\epsilon}$$

info@qedma.com

QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Error Correction is not enough, given specs & qubit numbers of near-term QPUs.

Error Mitigation (EM) boosts achievable circuit volumes.

Error mitigation toy model: depolarizing errors

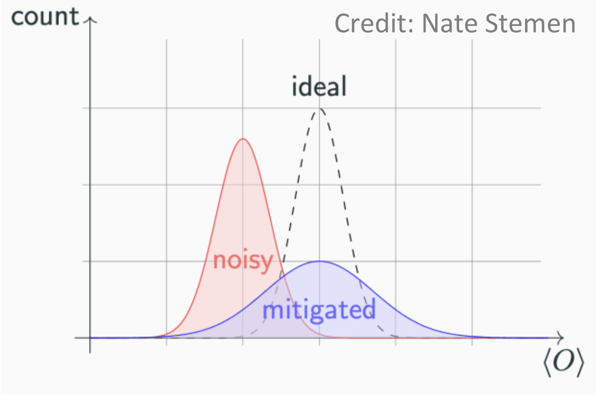
- Unmitigated observable: error-induced bias and quantum variance

$$\langle O(n)\rangle_{raw} = \langle O\rangle_{ideal} e^{-IF\cdot V} \pm \frac{1}{\sqrt{n}} \xrightarrow[\text{for accuracy } \epsilon]{\text{Requirements}} V_{raw} \leq \frac{\epsilon}{IF}, \quad n_{raw} \geq \frac{1}{\epsilon^2}$$

- Perfect mitigation: rescale to remove bias

$$\langle O\rangle_{miti} \equiv \langle O(n)\rangle_{raw} e^{IF\cdot V} = \langle O\rangle_{ideal} \pm \frac{e^{IF\cdot V}}{\sqrt{n}}$$

- Cost: exponential overhead $n_{miti} = \frac{e^{2IF\cdot V_{miti}}}{\epsilon^2}$



- Gain: dramatic circuit volume boost, even at moderate overhead x

$$V_{miti}(n_{miti} = x n_{raw}) = \frac{\log(\sqrt{x})}{IF} \longrightarrow \frac{V_{miti}}{V_{raw}} = \frac{\log(\sqrt{x})}{\epsilon}$$

info@qedma.com

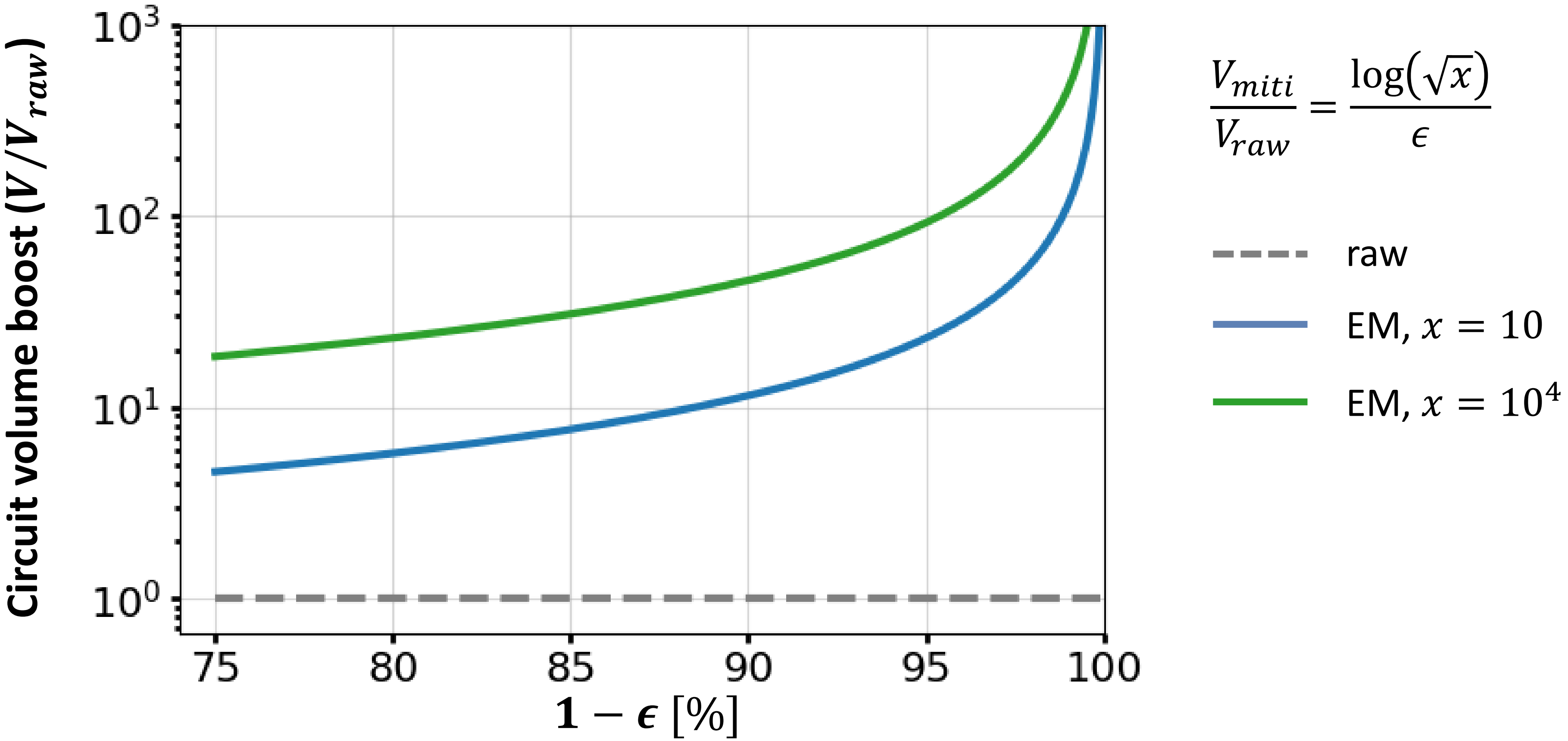
QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Error Correction is not enough, given specs & qubit numbers of near-term QPUs.

Error Mitigation (EM) boosts achievable circuit volumes.

Error mitigation delivers dramatic circuit volume boosts



High-accuracy EM enables to run up to x1000 larger quantum circuits!

ori.alberton@qedma.com

QEDMA

Errors severely limit the quantum circuit volumes executable on QPUs.

Error Correction is not enough, given specs & qubit numbers of near-term QPUs.

Error Mitigation (EM) boosts achievable circuit volumes.

EM will soon enable the first quantum advantages.

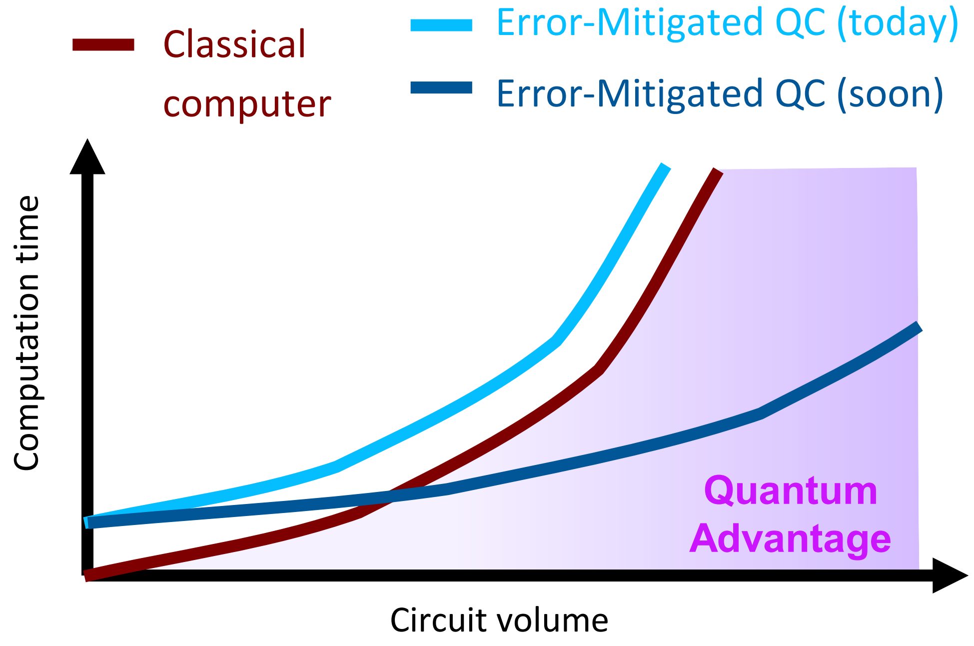
Can error mitigation enable quantum advantage?

Asymptotically: **polynomial quantum advantages** vs. classical simulation

EM scales mildly exponentially:
$T_Q = e^{\lambda_{EM} \cdot IF \cdot V}$

Examples: Cai et. al 2023; Bounds: Takagi et. al 2022, Quek et. al 2022, Schuster et. al 2024.

Classical computation scales wildly exponentially: $T_C = e^{\lambda_c V}$



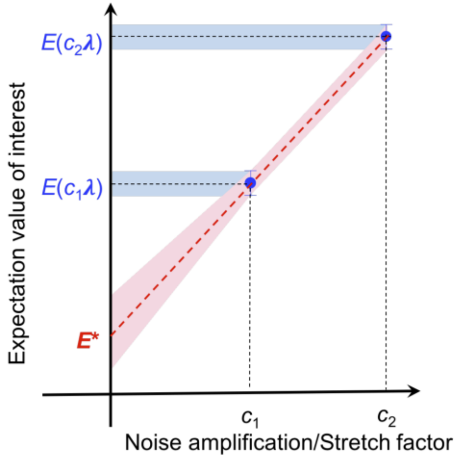
$T_c \propto T_Q^{O(1/IF) \sim 1000}$

Dramatic finite quantum advantage

info@qedma.com

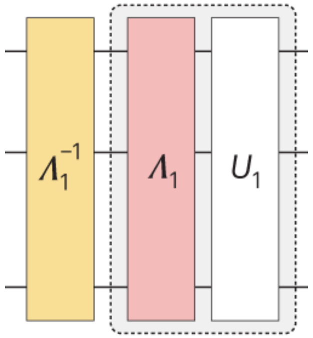
QESEM balances speed with reliability

Zero-Noise Extrapolation (ZNE):
amplify errors → extrapolate to the zero noise



Less overhead,
Heuristic

Probabilistic Error Cancellation (PEC):
locally invert errors with a
quasi-probabilistic circuit ensemble



Reliable,
Theoretical guarantees

QESEM:
efficient, scalable
quasi-probabilistic
mitigation

Clifford data regression:
infer from noisy Clifford circuits

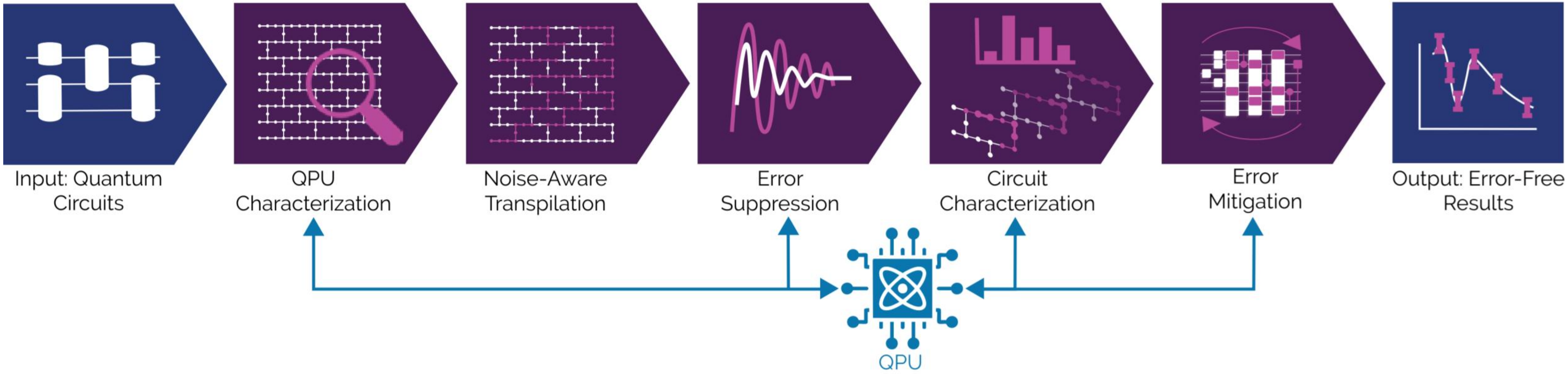
Virtual distillation:
measure entangled copies

Not covered in this talk

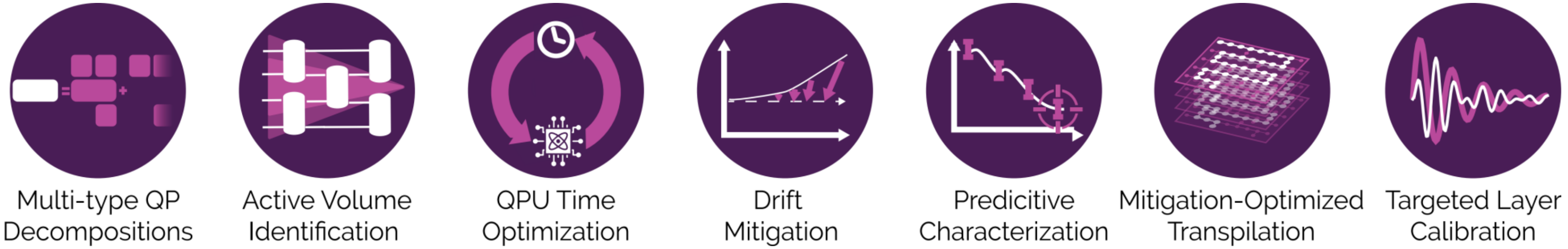
info@qedma.com

QESEM: efficient and reliable
Quantum Error Suppression & Error Mitigation software

QESEM Workflow



Key innovations



info@qedma.com

QESEM Client UX

Circuits and observables preparation

```
circuit = qiskit.QuantumCircuit(5)
circuit.cx(0, 1)
circuit.cx(2, 3)
circuit.cx(1, 2)
circuit.cx(3, 4)
avg_magnetization = qiskit.quantum_info.SparsePauliOp.from_sparse_list(
    [("Z", [q], 1 / 5) for q in range(5)], num_qubits=5
)
other_observable = qiskit.quantum_info.SparsePauliOp.from_sparse_list(
    [("ZZ", [0, 1], 1.0), ("XZ", [1, 4], 0.5)], num_qubits=5
)
```

QESEM job creation

```
job = qedma_client.create_job(
    circuit = circuit,
    observables = [avg_magnetization, other_observable],
    observables_metadata = [
        qedma_api.ObservableMetadata(description = "avg_magnetization"),
        qedma_api.ObservableMetadata(description = "other_observable")
    ],
    empirical_time_estimation = True,
    precision = 0.1,
    backend = "fake_fez",
    description = "my qesem job",
)
2025-07-01 14:32:11.027 | INFO     | Submitting new job
2025-07-01 14:32:12.748 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] New job created
```

Obtain time estimation

```
qedma_client.wait_for_time_estimation(job_id = job.job_id)
2025-07-01 14:36:01.736 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] Time estimation: [5.0 minutes]
```

QESEM job execution initiation

```
qedma_client.start_job(
    job_id = job.job_id,
    max_qpu_time = datetime.timedelta(minutes = 5)
)
2025-07-01 14:36:11.772 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] Starting job
```

Monitor progress intermediate results final results

```
job = qedma_client.wait_for_job_complete(job_id = job.job_id)
2025-07-01 14:36:18.124 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Queued]
2025-07-01 14:36:18.125 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Initial Transpilation]
2025-07-01 14:36:18.126 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Qubit Selection]
2025-07-01 14:37:01.916 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Calibration]
2025-07-01 14:37:01.918 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Characterization]
2025-07-01 14:38:17.658 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] step: [Batch Mitigation #1]
2025-07-01 14:39:54.997 | INFO     | [60ff19853e944112a44e655a5ecb2fb0] Final results: [[<SparseObservable with 5 terms on 5 qubits:
    (0.2+0j)(Z_0) + (0.2+0j)(Z_1) + (0.2+0j)(Z_2) + (0.2+0j)(Z_3) + (0.2+0j)(Z_4)>: (1.000676575536313 ± 0.0033883682697840594),
    <SparseObservable with 2 terms on 5 qubits: (1+0j)(Z_1 Z_0) + (0.5+0j)(Z_4 X_1)>: (1.009224816958378 ± 0.0020171731955879392)]]
```

QESEM Execution Metrics (available to end-users)

Noisy results (i.e. without QESEM)

```python
job = qedma_client.get_job(job_id = "<job_id>", include_results = True)
for observable, expectation_value in job.results:
    noisy_value = [noisy_val for o, noisy_val in job.noisy_results if o==observable][0]
    print(f"Observable: {observable}")
    print(f"Mitigated value: {expectation_value}")
    print(f"Noisy value: {noisy_value}")
```

```
Observable: {'X0,X1': 0.0058, 'X0,X1,X2,X3': 0.0124}
Mitigated value: -0.6426069139973665 ± 0.06942964571907104
Noisy value: -0.445530371314418 ± 0.03295058637115513
```

Shot count

```python
print(f"Total shots: {job.execution_details.total_shots}")
print(f"Mitigation shots: {job.execution_details.mitigation_shots}")
```

```
Total shots: 1259210
Mitigation shots: 253210
```

Total QPU time

```python
print(f"QPU time: {job.qpu_time['execution']}")
```

```
QPU time: 0:00:08
```

Gate fidelities

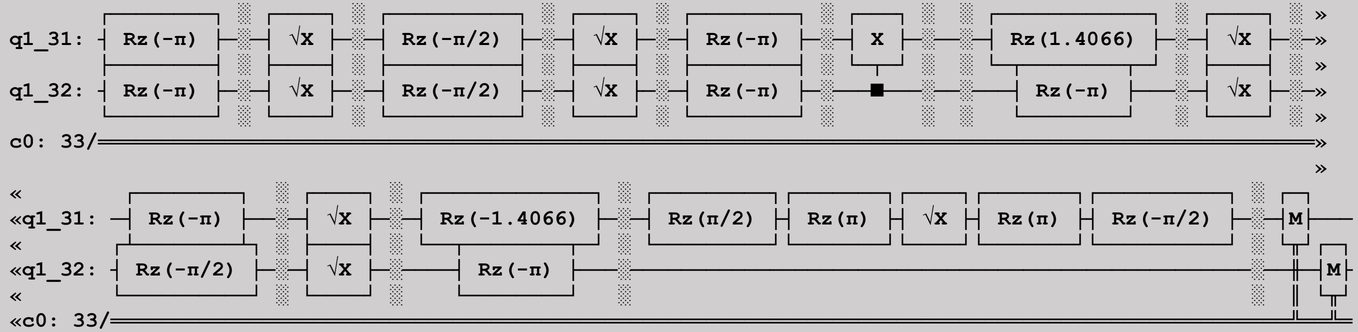
```python
print(f"Gates fidelities: {job.execution_details.gate_fidelities}")
```

```
Gate fidelities:  {'CNOT': 0.9900899581083492, 'ID1Q': 0.9989167673500885}
```

Transpiled circuit

```python
print(f"Transpiled circuit:")
print(job.execution_details.transpiled_circuits[0].circuit.draw(output="text", idle_wires=False, fold=10,vertical_compression='high'))
print(f"Transpiled circuit qubit mapping: {job.execution_details.transpiled_circuits[0].qubit_map}")
print(f"Transpiled circuit number of measurement bases: {job.execution_details.transpiled_circuits[0].num_measurement_bases}")
```

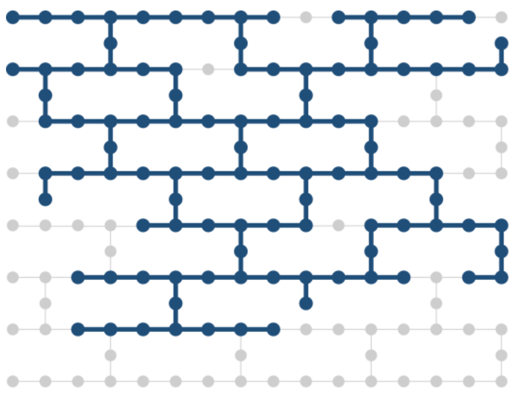
```
Transpiled circuit:
```



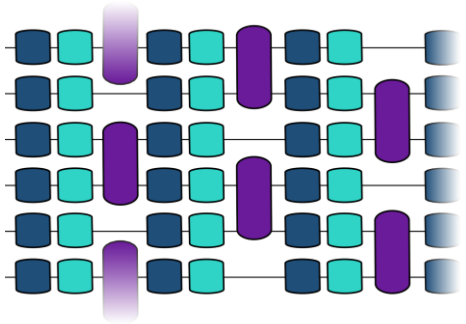

```
Gates fidelities: {'CNOT': 0.989021363408563}
Transpiled circuit qubit mapping: {0: 32, 1: 31}
Transpiled circuit number of measurement bases: 1
```

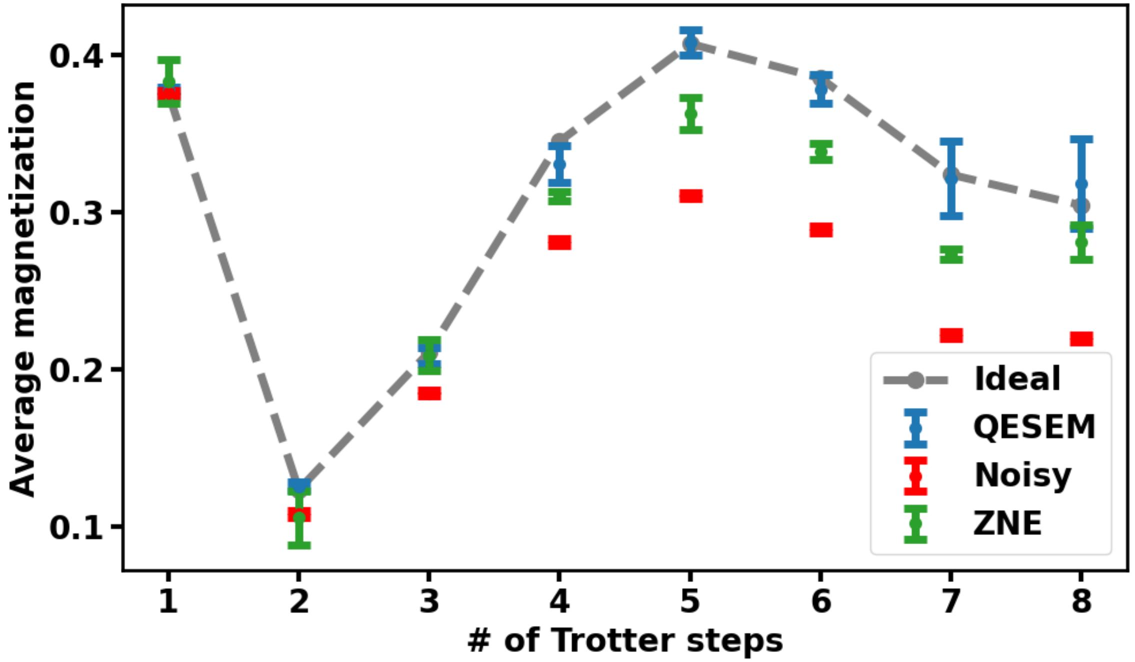
QEDMA

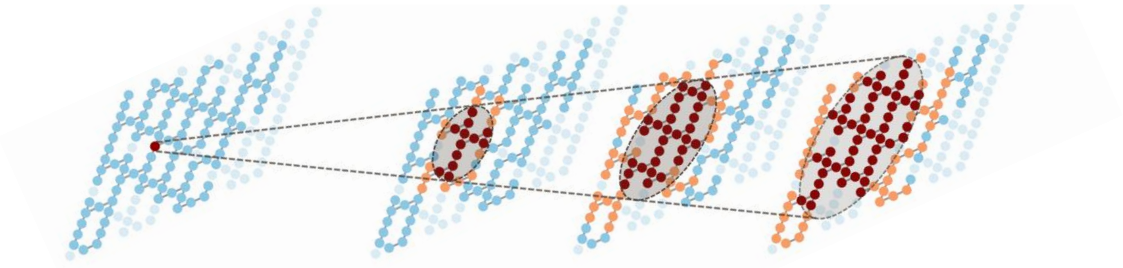
Demo: approaching quantum advantage at high accuracy



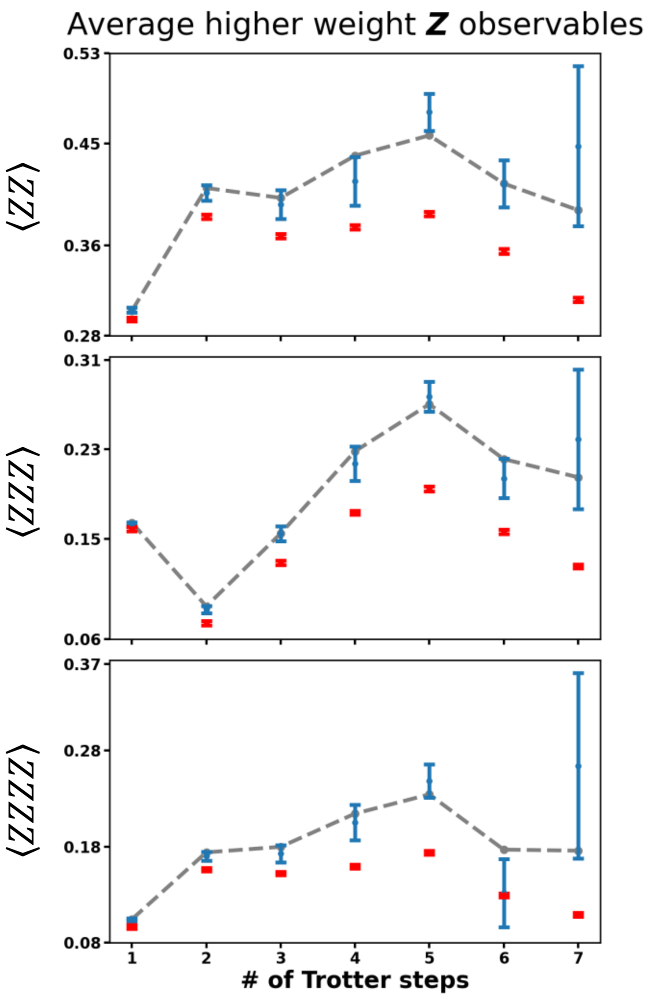
2D Trotter-Ising on **100** qubits with fractional ZZ angle (IBM Marrakesh)







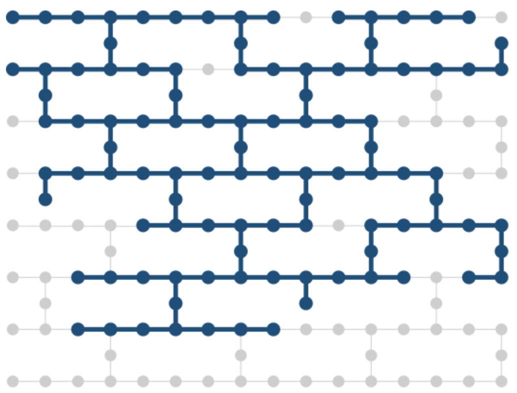
Total volume: 832 R_{zz} gates (1,664 CZ gates)
Active volume: 301 R_{ZZ} gates (602 CZ gates)
Base of lightcone: 54 qubits



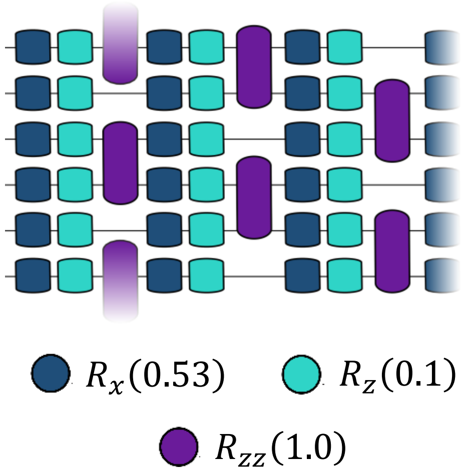
info@qedma.com

QEDMA

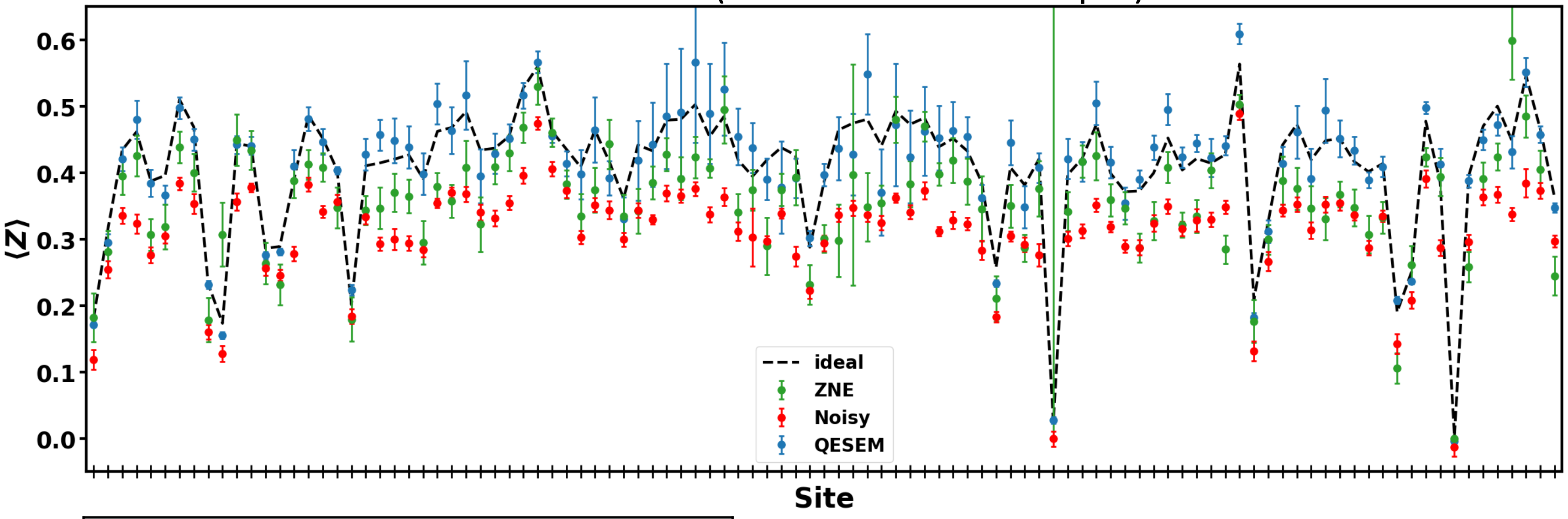
Unbiased mitigation of all 100 qubits with QESEM

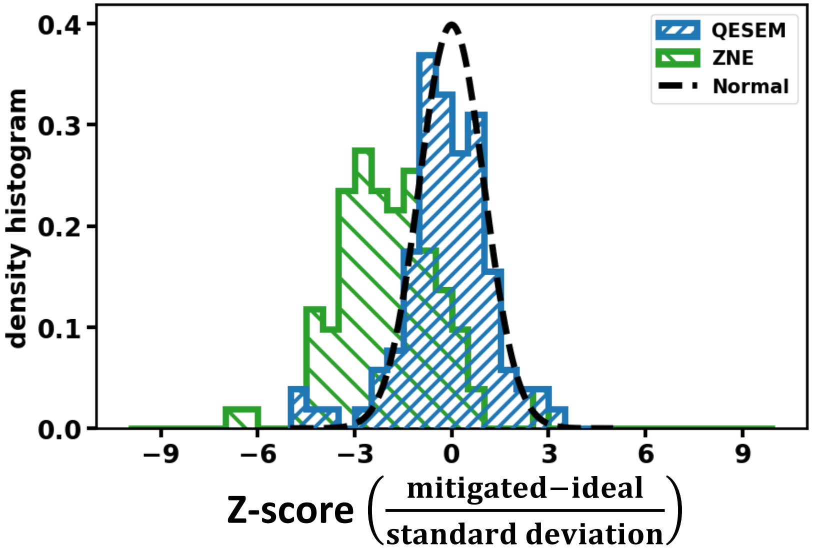


2D Trotter-Ising on **100** qubits with fractional ZZ angle (IBM Marrakesh)



All Z observables (focus on Trotter step 5)

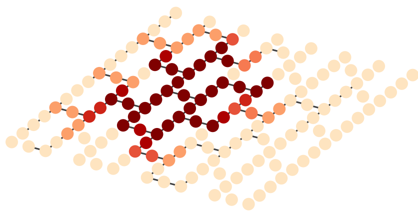




Median Z-score

QESEM: -0.15

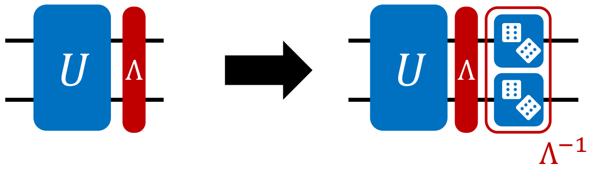
ZNE: -2.08



info@qedma.com

Focus: multi-type quasi-probability (QP) decompositions

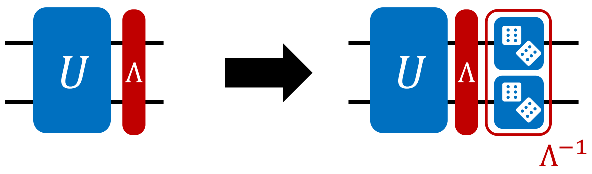
- Standard PEC:
  - **Assume errors occur only after** (or only before) the noisy operation
  - **Insert correction gates only after** (or only before) the noisy operation
  - Incompatible or suboptimal for non-Clifford gates



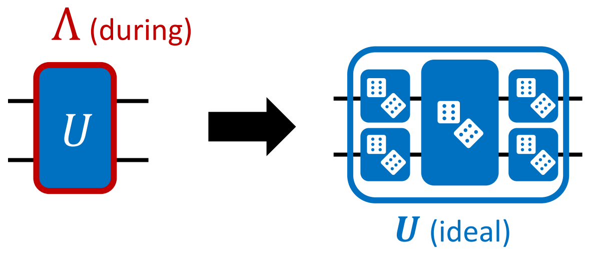
info@qedma.com

Focus: multi-type quasi-probability (QP) decompositions

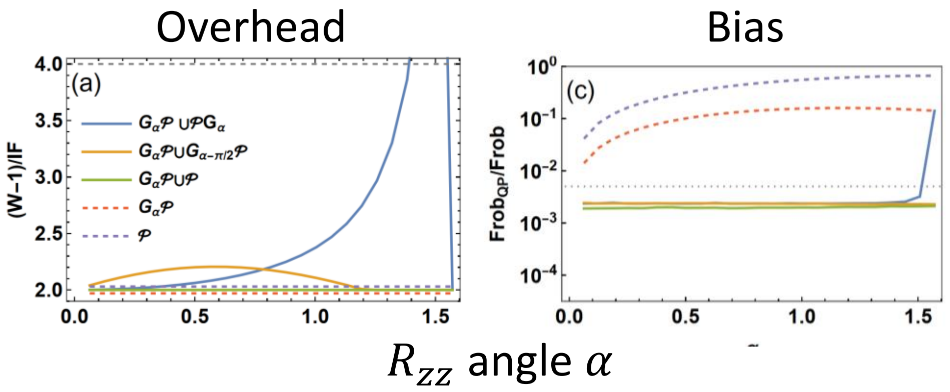
- Standard PEC:
  - **Assume errors occur only after** (or only before) the noisy operation
  - **Insert correction gates only after** (or only before) the noisy operation
  - Incompatible or suboptimal for non-Clifford gates



- Multi-type QP decompositions:
  - Accommodate also **errors during** the noisy operation
  - **Insert corrections before, after, and instead** of the noisy operation
  - Optimal mitigation for layers of general **non-Clifford 2q gates**
    ⇒ **reduces circuit depth by up to a factor of 2**



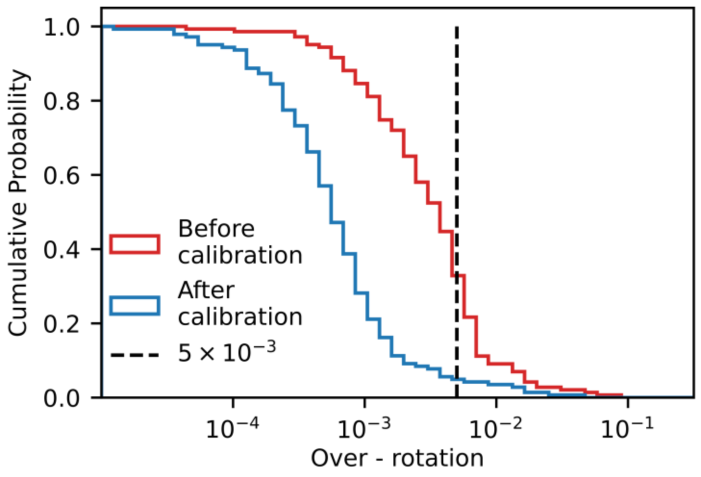
Multi-type QP decompositions optimize both overhead and bias



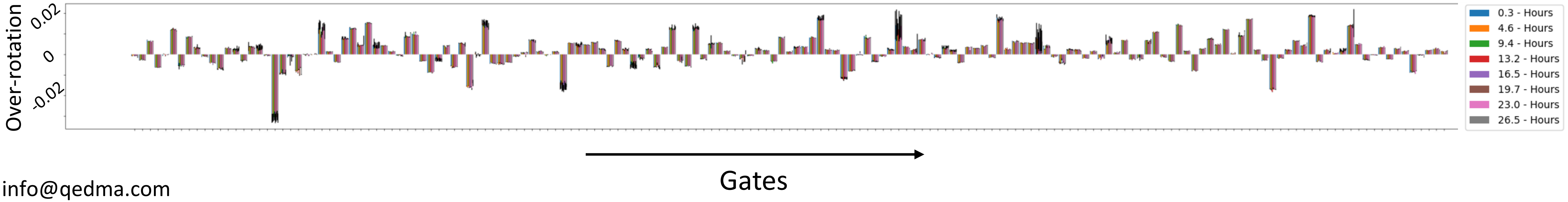
info@qedma.com

Focus: error suppression & calibration

- **Partial Pauli twirling** shapes non-ZZ coherent errors
- **Targeted calibration** eliminates over-rotation errors (active ZZ) in a crosstalk-resilient manner
- To reduce infidelity, we suppress additional coherent errors
- On Herons, native R_{ZZ} calibration is quite stable, so one calibration per QESEM run suffices



R_{zz} over-rotation stability

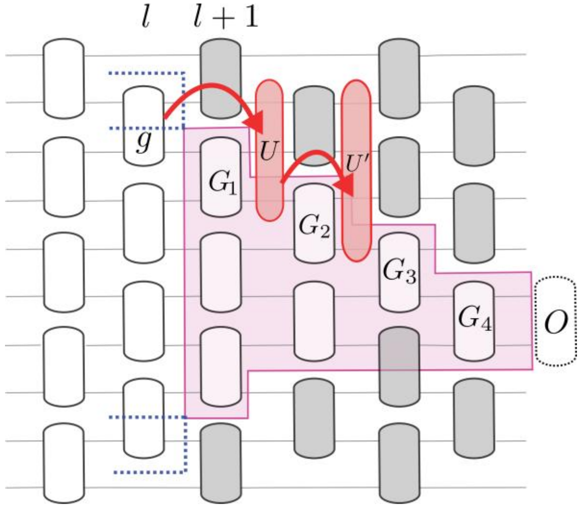


info@qedma.com

Focus: active volume identification

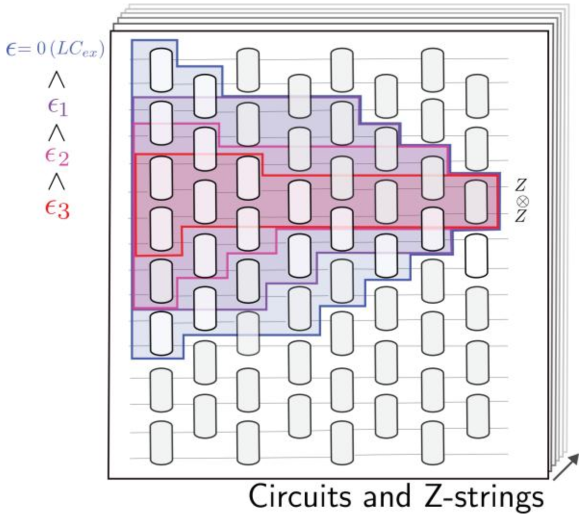
Basic idea: construct a separate lightcone (LC) for each constituent Pauli string in post-processing.

Candidate LC construction:
iteratively push-forward gates out of the LC
(with approximation parameter ϵ)

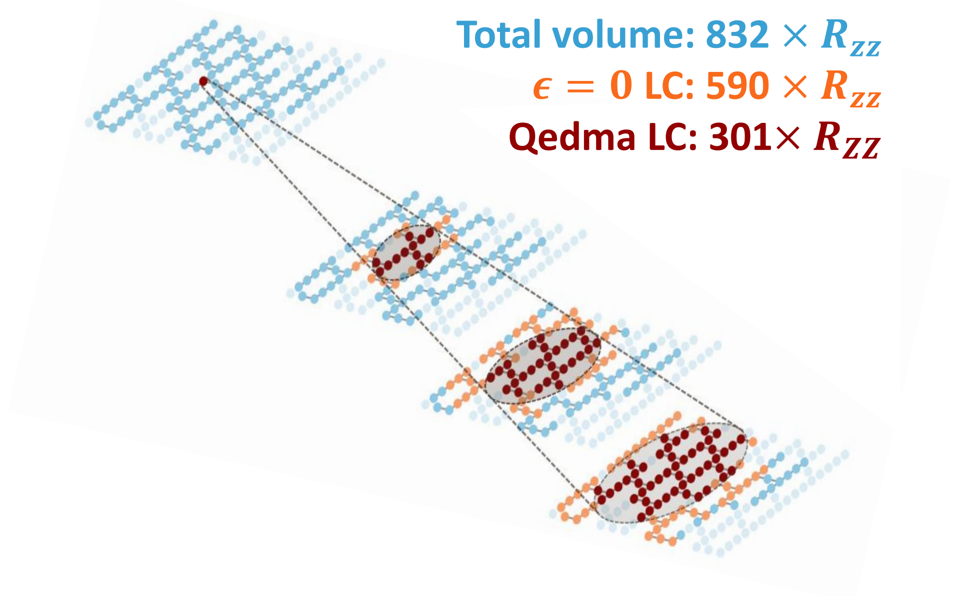


$$\|U - I \otimes B\| \leq \epsilon$$

LC selection:
empirically estimate the bias of each candidate
from executed mitigation circuits



Results:
Significantly smaller volume
necessitating mitigation

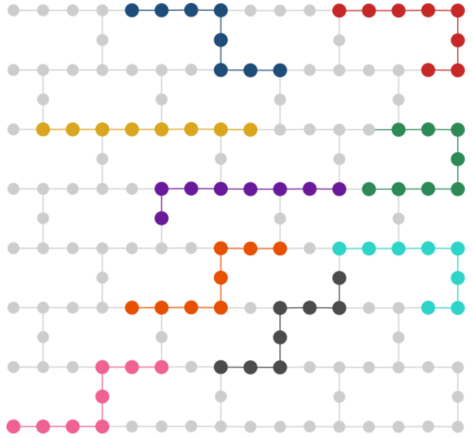


Total volume: 832 $\times R_{ZZ}$
$\epsilon = 0$ LC: 590 $\times R_{ZZ}$
Qedma LC: 301$\times R_{ZZ}$

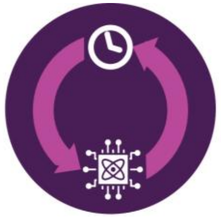
info@qedma.com

For a different approach see Eddins et al, 2409.04401

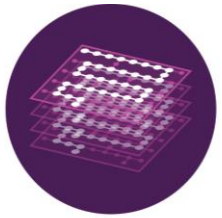
Demo: molecular VQE for the ground state of H_2O



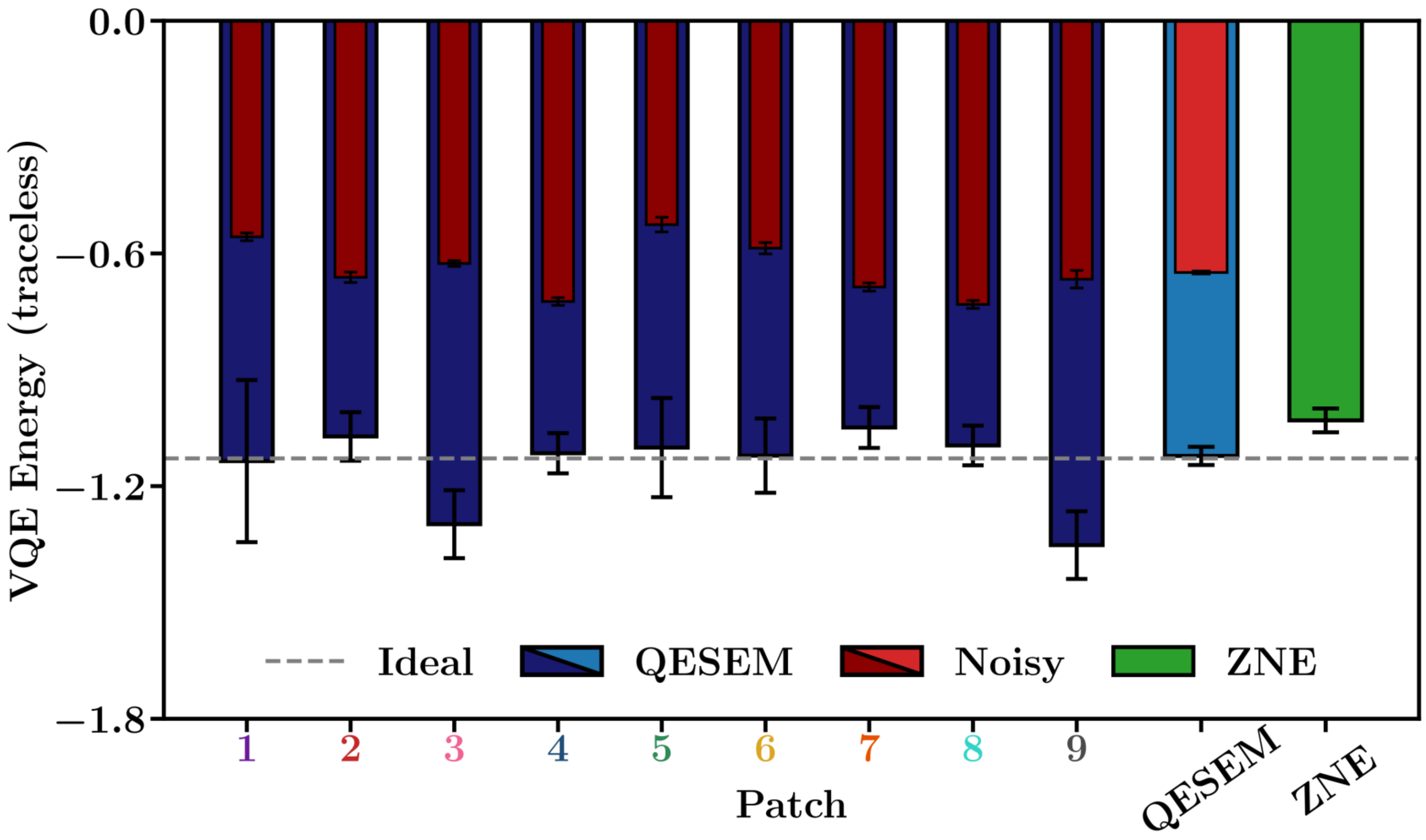
Parallelized molecular VQE (IBM Marrakesh).



65
meas. bases



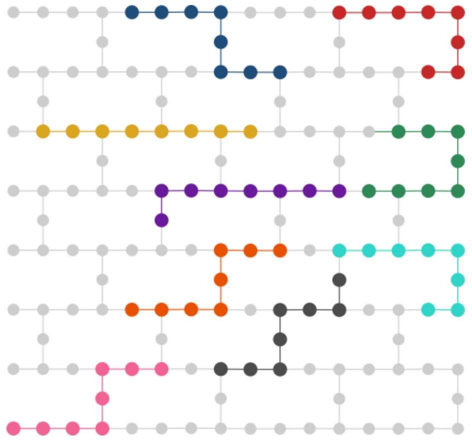
9 parallel
patches



info@qedma.com

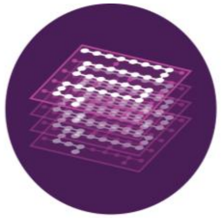
QEDMA

Demo: molecular VQE for the ground state of H_2O

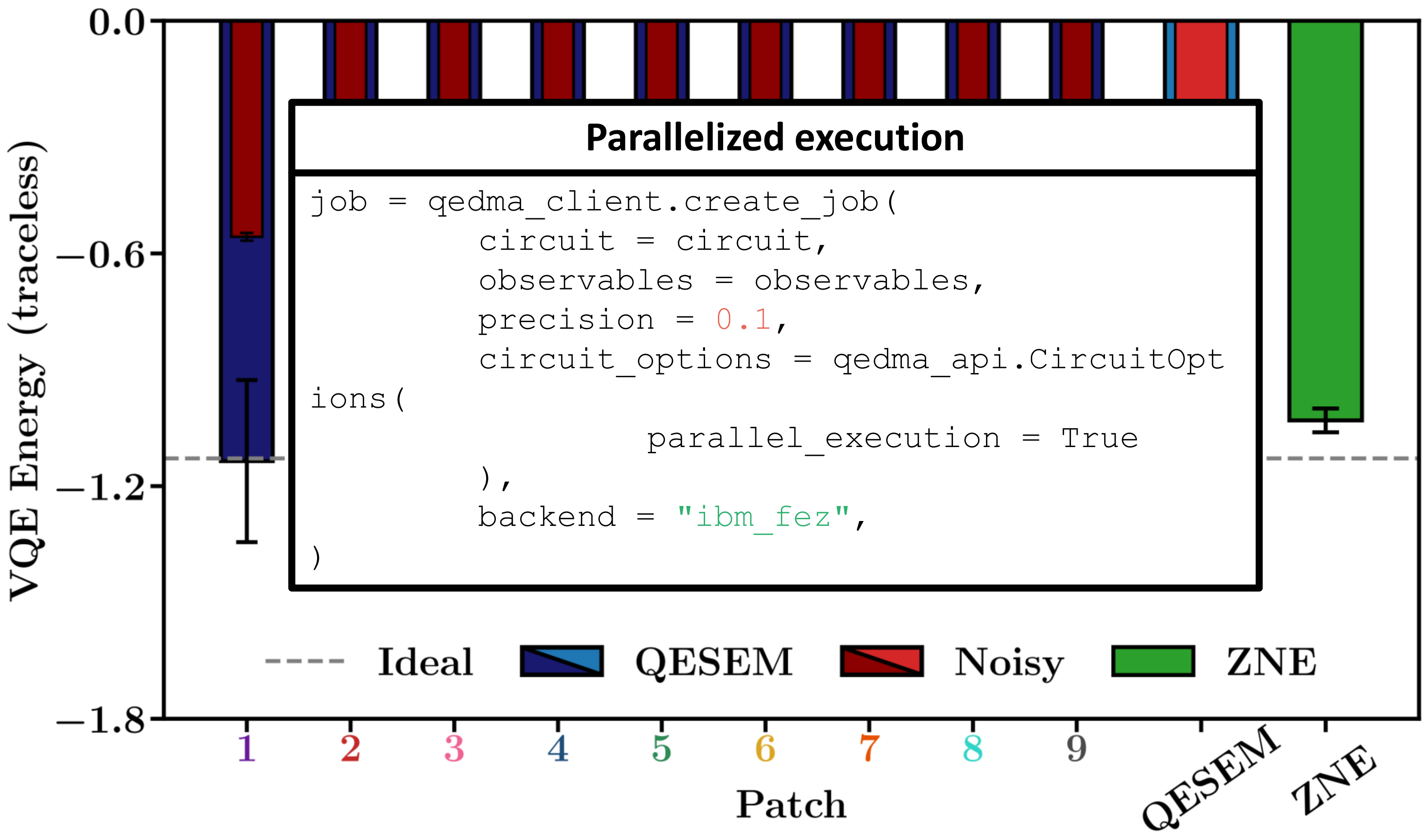


Parallelized molecular VQE (IBM Marrakesh).

65 meas. bases



9 parallel patches



Parallelized execution

```
job = qedma_client.create_job(
        circuit = circuit,
        observables = observables,
        precision = 0.1,
        circuit_options = qedma_api.CircuitOpt
ions(
                parallel_execution = True
        ),
        backend = "ibm_fez",
)
```

info@qedma.com

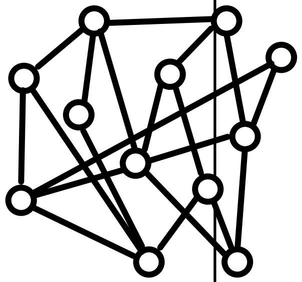
QESEM + IonQ + Amazon Braket: computing molecular energies on trapped ions

Application
VQE, O_2 molecule

Width
12 qubits

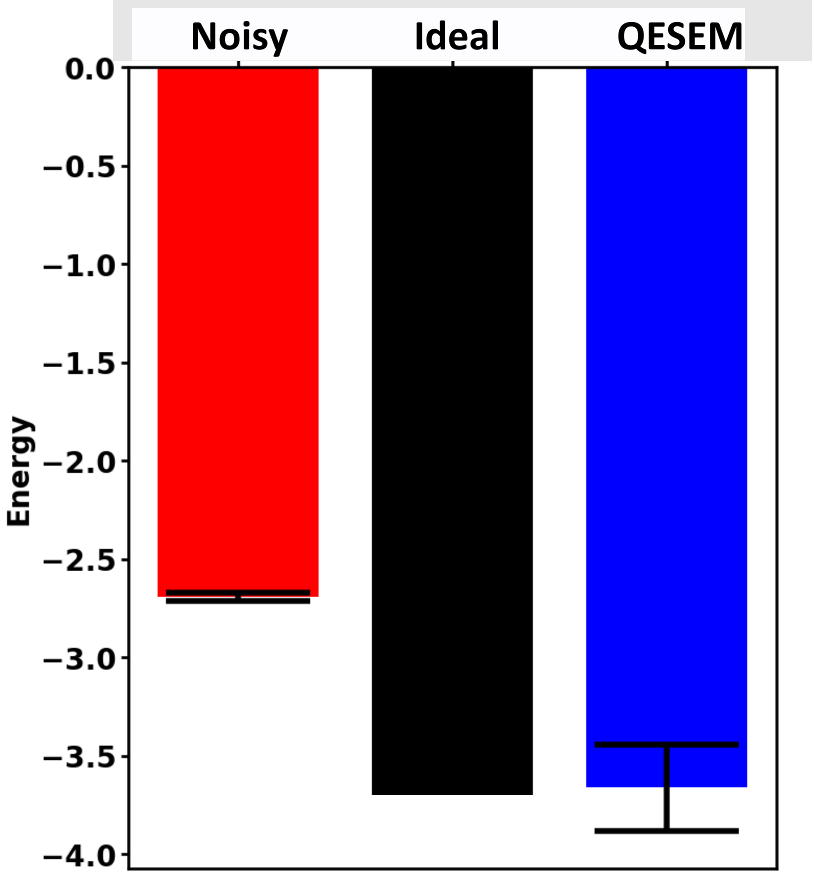
Volume
99 2-qubit gates
(20 distinct)

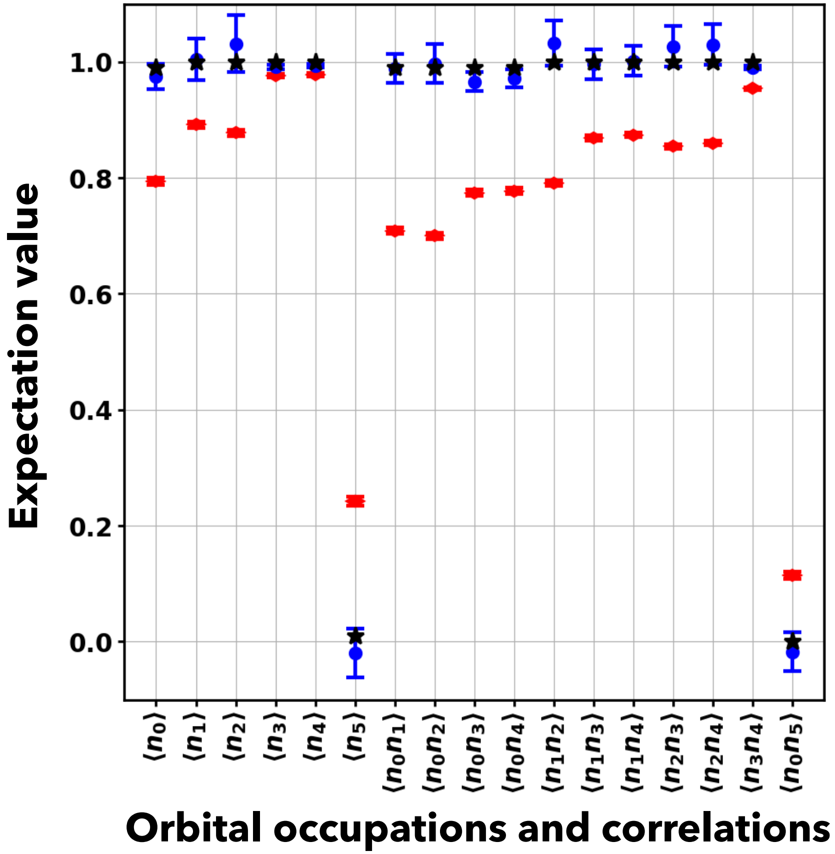
Connectivity
Sparse graph



Quantity
Energy, orbital occupation and correlations

- **Largest unbiased EM experiment on trapped ions.**
- **Largest VQE circuit with accurate output.**





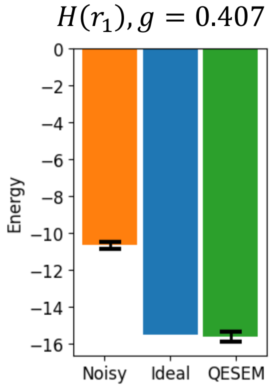
info@qedma.com

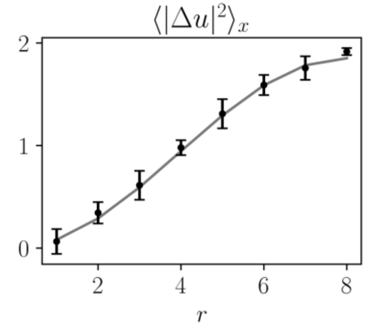
More details: aws.amazon.com/blogs/quantum-computing/…

QESEM in Action: Real-World Success Stories
Available via Qedma Cloud, On Premises or as a Qiskit Function

DESY – CQTA (Karl Jansen)

- Studying running couplings in U(1) lattice gauge theory in 2+1 dimensions with VQE (Paulo Itaborai, Arianna Crippa)
- Measuring moments of velocity fields in VQA based approach for fluid-dynamics (Miriam Goldack)





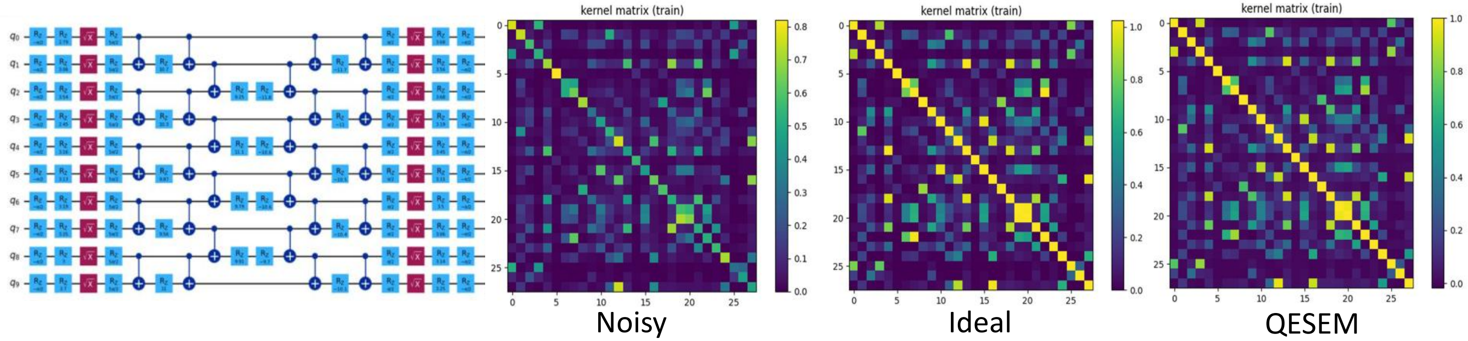
Pacific Northwest National Laboratory, (Muqing Zheng)

"Coupled Cluster Downfolding Theory in Simulations of Chemical Systems on Quantum Hardware", arxiv:2507.01199

Softbank (Hiroshi Yamauchi)

QESEM applied to ensembles of circuits in a Quantum Kernel Method

Based on *Parametrized energy-efficient quantum for network fault diagnosis,* IEEE 2024 (arxiv:2405.09724)



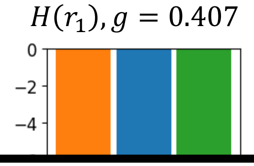
info@qedma.com

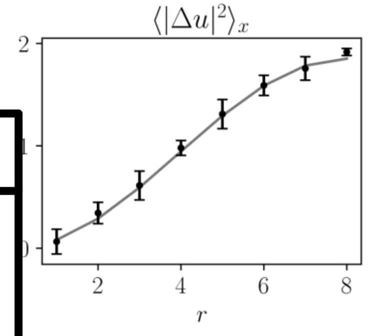
QEDMA

QESEM in Action: Real-World Success Stories
Available via Qedma Cloud, On Premises or as a Qiskit Function

DESY – CQTA (Karl Jansen)

DESY. QUANTUM — Center for Quantum Technology and Applications





- Studying running couplings in U(1) lattice gauge theory in 2+1 di
(Paulo Itaborai, A
- Measuring mom
approach for flui

Pacific Northwest

"Coupled Cluster Do
on Quantum Hardw

Softbank (Hiroshi Y

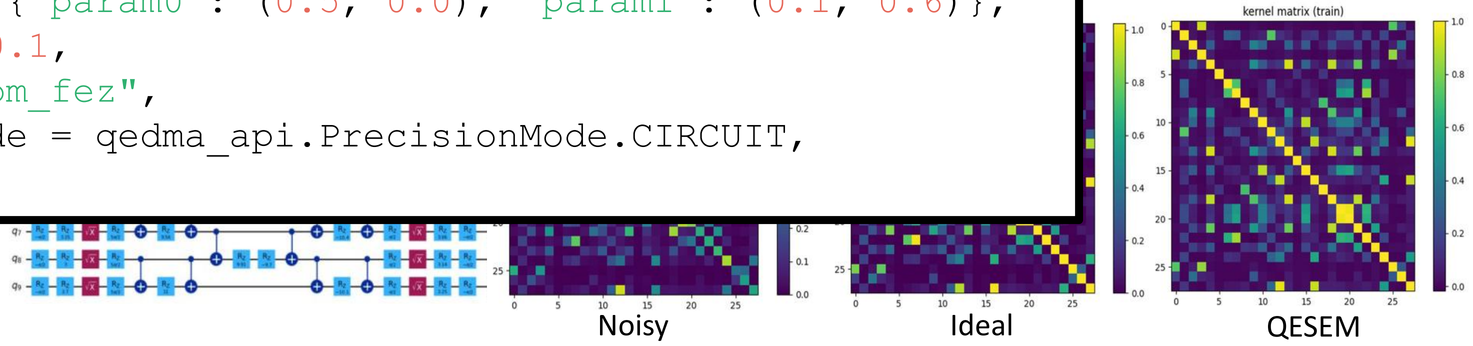
QESEM applied to e
circuits in a Quantu

Based on Parametrized e
for network fault diagnos

Consolidated mitigation of multiple (parameterized) circuits

```
circ = qiskit.QuantumCircuit(4)
circ.x(3)
for i in range(2):
        circ.rx(qiskit.circuit.Parameter(f"param{i}"), i)
circ.cx(0, 1)
circ.cx(2, 3)

qedma_client.create_job(
        circuit = circ,
        observables = (
                qedma_api.Observable({"X1": 1.0, "Z0,Z3": 0.3}),
                qedma_api.Observable({"X2": 1.0, "Z0,Z3": 0.3}) ),
        parameters = {"param0": (0.5, 0.0), "param1": (0.1, 0.6)},
        precision = 0.1,
        backend = "ibm_fez",
        precision_mode = qedma_api.PrecisionMode.CIRCUIT,
)
```

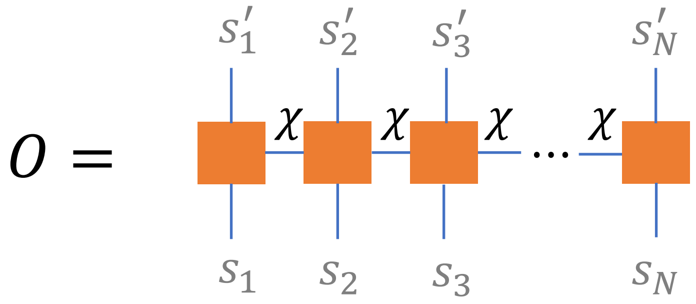


info@qedma.com

Quantum-enhanced MPO computation

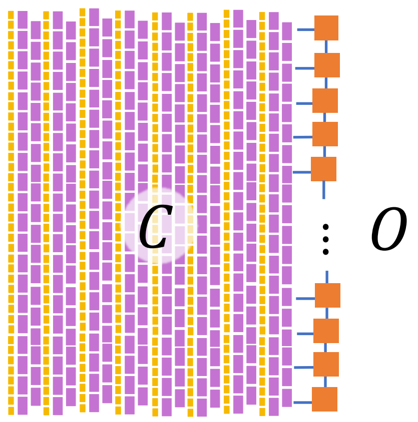
MPO computation: efficiently compressing information

- General operator on N qubits is described by 4^N parameters
  → computation scales exponentially with the number of qubits
- MPOs: an efficient representation of an operator O as a product of matrices with bond dimension χ
  → computation is efficient: scales as $\chi^3 \cdot N$
- Widely used in quantum many-body physics, quantum chemistry and machine learning



Quantum-enhanced MPO computation

New method for application of a quantum circuit C to an MPO by measuring $\langle 0|C^{+}OC|0\rangle$ on a quantum computer with QESEM.



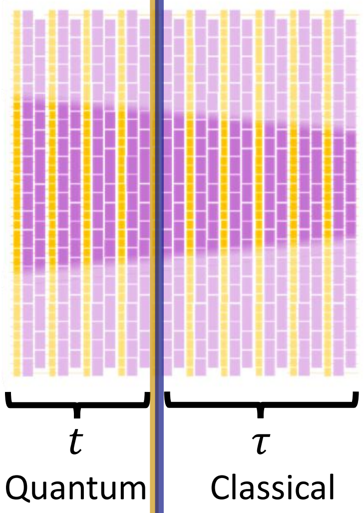
Accelerating the quantum advantage timeline with HPC-QPU integration

Hybrid QPU-HPC circuit execution:

$$\langle\psi(T)|O|\psi(T)\rangle = \langle\psi(t)|\ O(\tau)|\psi(t)\rangle$$

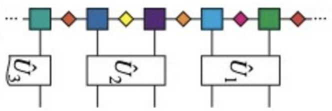
$|\psi(\tau)\rangle$

+



$\langle O\rangle$

$O(\tau)$

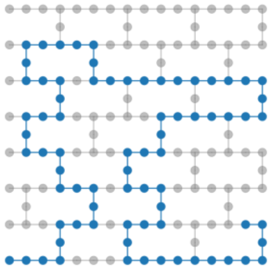


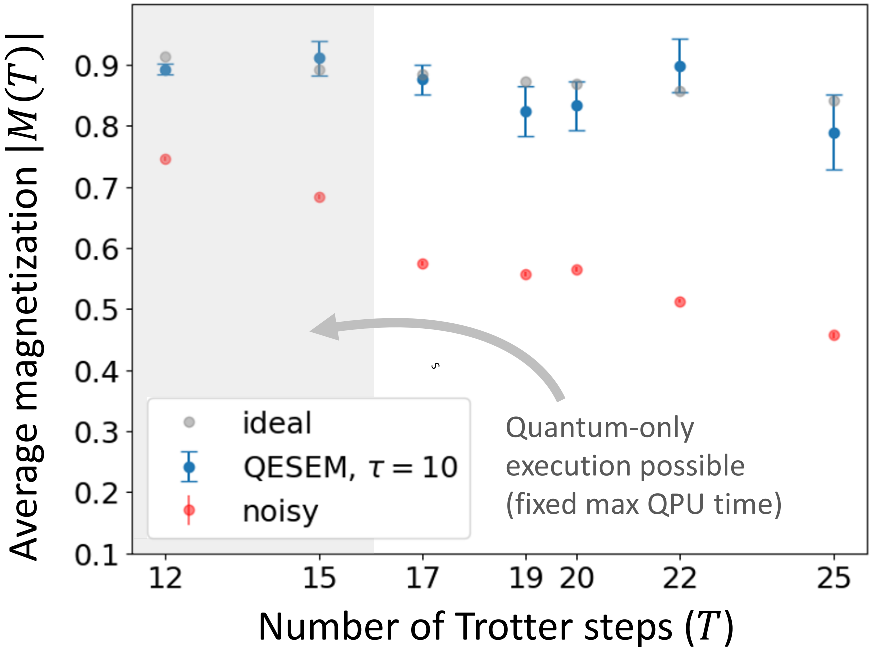
Yunoki group

Total evolution time:

$$T = t + \tau$$

Demonstration:
kicked Ising model
(75 qubits)





info@qedma.com

QEDMA

Hybrid QPU + HPC circuit execution with QESEM:

Significant boost to the scale of quantum simulations

Lower fidelity requirement for quantum advantage

New functionalities for quantum processing of tensor-network data structures

Roadmap for quantum-enhanced HPC simulations with QESEM

Estimated computation time

For $\langle\psi(t)|O_{MPO}|\psi(t)\rangle$ given MPO with bond dimension $\chi_{in} = 10^4$

Hardware	t=5 Runtime [hr]		t=10 Runtime [hr]		t=50 Runtime [hr]	
	C	Q	C	Q	C	Q
Nvidia Grace Hopper*	276	-	3325	-	Unfeasible	
IBM Heron r2 (F=99.5%)	17	0.5	17	3	Unfeasible	
IBM (F=99.9%)	17	~0	17	0.2	17	3

* Estimation from: $\chi(t) = \chi_{in} e^{3vt}$ and $T_{classical} \sim \alpha \chi_{in}^3 \sum_{n=1}^{t} e^{3vn}$,
where $\alpha_{GPU} \approx 1.7 \times 10^{-11}$ estimated from Haghshenas et al, arXiv:2503.20870

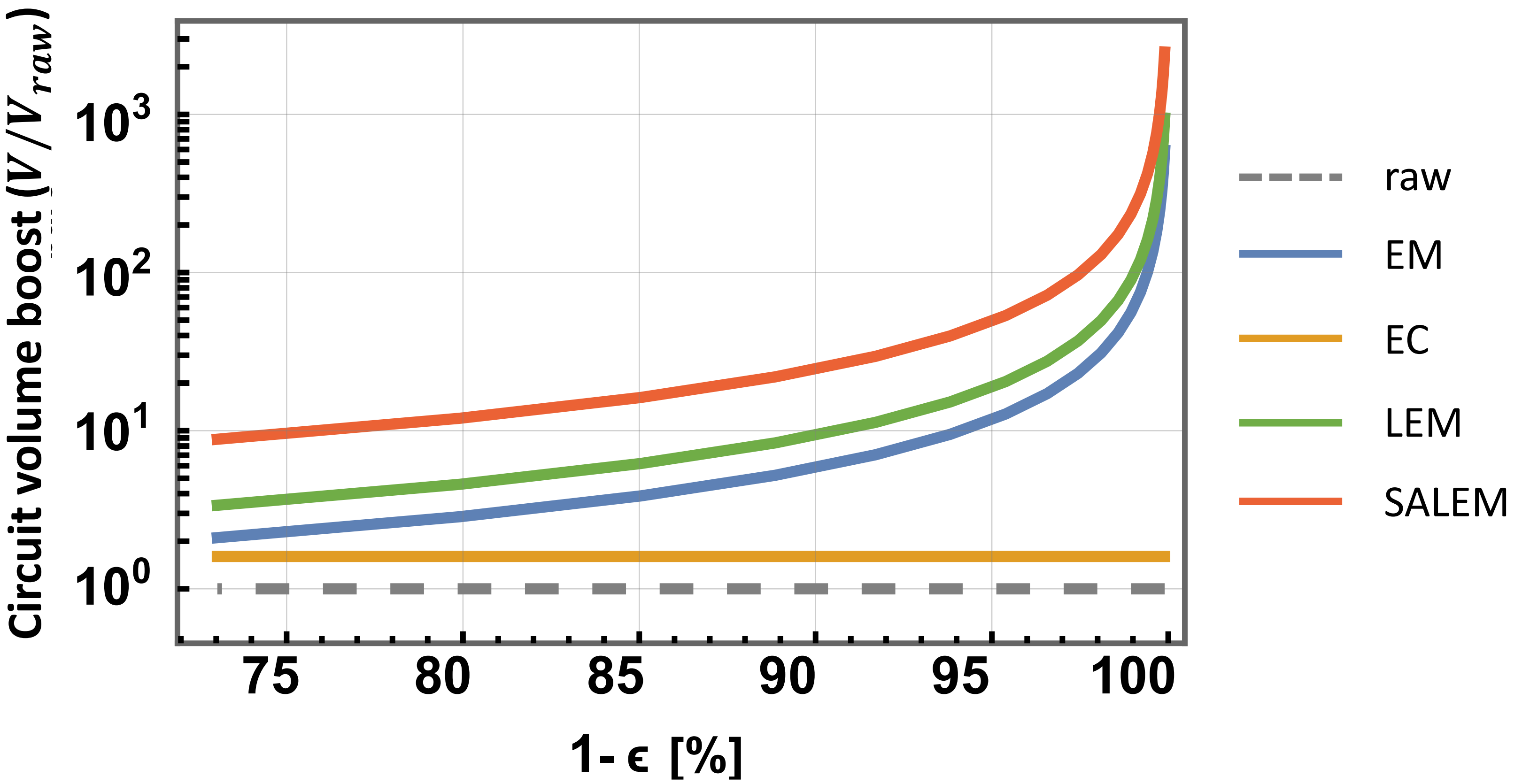
QESEM + HPC → quantum-enhanced classical simulations today

+

>

info@qedma.com

EM is here to stay: combining EM+EC for scalable QC



SALEM enables Megaquop

Example: BB qLDPC code

[[n,k,d]] = [[144, 12, 12]]

$$V_{EC} = 6 \cdot 10^5$$

$$V_{SALEM} = 6 \cdot (10^7 - 10^8)$$

Assumes:

Required precision $\epsilon = 0.01$

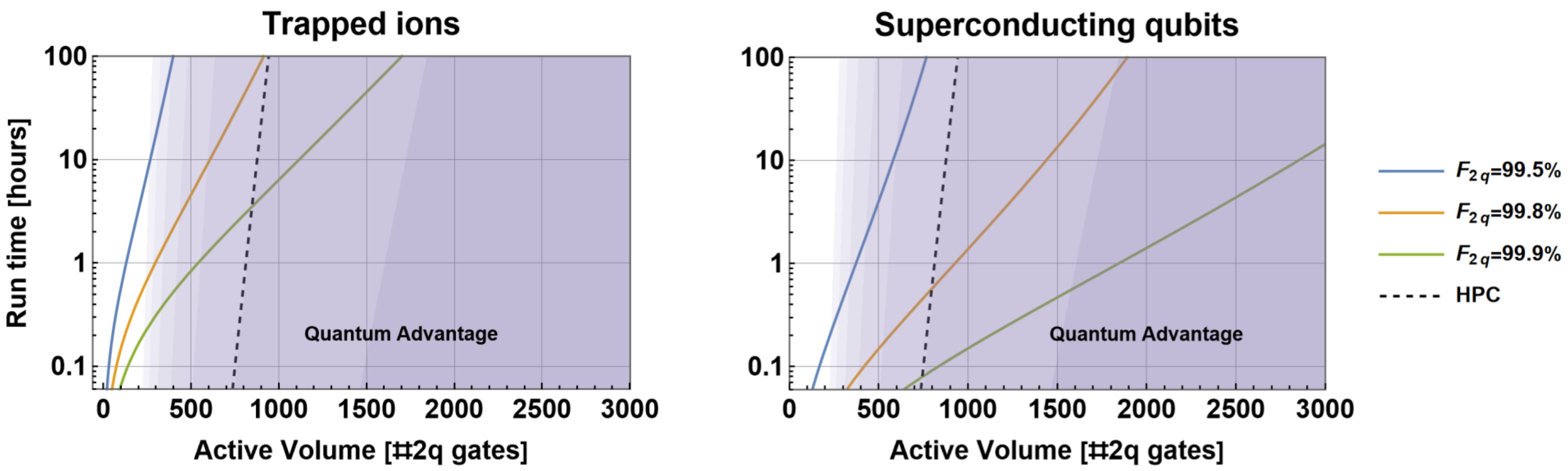
Physical IF=1×10^{-3}

Allowed overhead for EM = 100

SALEM described in upcoming paper: “*Syndrome-aware mitigation of logical errors*”

info@qedma.com

Error mitigation will soon unlock quantum advantage



Check out our papers:

- *On the importance of error mitigation for quantum computation* (arXiv: 2503.17243)
- *Reliable, high-accuracy error mitigation for utility-scale quantum circuits (arXiv: 2508.10997)*
- *Syndrome-aware mitigation of logical errors (coming soon)*

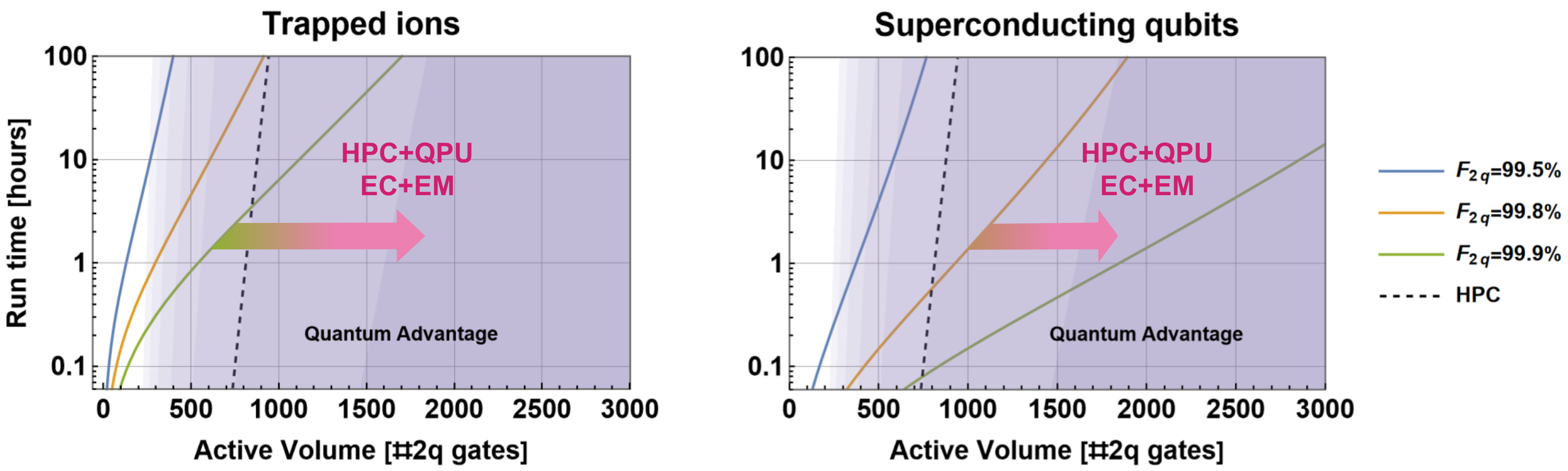
Reach out!

netanel.Lindner@qedma.com

info@qedma.com

qedma.com

Error mitigation will soon unlock quantum advantage



EM integrated with HPCs and error correction accelerates the quantum advantage timeline

Check out our papers:

- *On the importance of error mitigation for quantum computation* (arXiv: 2503.17243)
- *Reliable, high-accuracy error mitigation for utility-scale quantum circuits (arXiv: 2508.10997)*
- *Syndrome-aware mitigation of logical errors (coming soon)*

Reach out!

netanel.lindner@qedma.com

info@qedma.com

qedma.com