Learning Junta Distributions, Quantum Junta States, and QAC^0 Circuits

Jinge Bao[1,2] Francisco Escudero Gutiérrez[3]

[1]U of Edinburgh [2]CQT, NUS [3]QuSoft & CWI

ICML 2026 • Seoul

The setting: learning from few samples

A core question in learning theory:

If we know an unknown object has ***structure****, how efficiently can we learn it?*

The setting: learning from few samples

A core question in learning theory:

*If we know an unknown object has **structure**, how efficiently can we learn it?*

We consider three structured families on n bits/qubits:

- **Junta distributions** $p : \{-1, 1\}^n \to [0, 1]$ depends on only k bits

The setting: learning from few samples

A core question in learning theory:

*If we know an unknown object has **structure**, how efficiently can we learn it?*

We consider three structured families on n bits/qubits:

- **Junta distributions** $p : \{-1, 1\}^n \to [0, 1]$ depends on only k bits
- **Quantum junta states** a state on k qubits, maximally mixed on the rest (new)

The setting: learning from few samples

A core question in learning theory:

*If we know an unknown object has **structure**, how efficiently can we learn it?*

We consider three structured families on n bits/qubits:

- **Junta distributions** $p : \{-1, 1\}^n \to [0, 1]$ depends on only k bits
- **Quantum junta states** a state on k qubits, maximally mixed on the rest (new)
- QAC^0 **circuits** constant-depth quantum circuits with Toffoli gates

The setting: learning from few samples

A core question in learning theory:

*If we know an unknown object has **structure**, how efficiently can we learn it?*

We consider three structured families on n bits/qubits:

- **Junta distributions** $p : \{-1, 1\}^n \to [0, 1]$ depends on only k bits
- **Quantum junta states** a state on k qubits, maximally mixed on the rest (new)
- QAC^0 **circuits** constant-depth quantum circuits with Toffoli gates

One algorithmic idea handles all three.

The unifying message

Key observation

A function / state is **close to a k-junta** when its Fourier / Pauli spectrum is

low-degree concentrated ***and*** **sparse.**

The unifying message

Key observation

A function / state is **close to a k-junta** when its Fourier / Pauli spectrum is

low-degree concentrated ***and*** **sparse.**

- Juntas $\Rightarrow$ both properties hold by definition.

The unifying message

Key observation

A function / state is **close to a k-junta** when its Fourier / Pauli spectrum is

low-degree concentrated ***and*** **sparse.**

- Juntas $\Rightarrow$ both properties hold by definition.
- QAC^0 Choi states $\Rightarrow$ also satisfy both (our new contribution).

The unifying message

Key observation

A function / state is **close to a k-junta** when its Fourier / Pauli spectrum is

low-degree concentrated ***and*** **sparse.**

- Juntas $\Rightarrow$ both properties hold by definition.
- QAC^0 Choi states $\Rightarrow$ also satisfy both (our new contribution).

The algorithm: estimate the few low-degree, sparse Fourier / Pauli coefficients, then truncate.

That single recipe gives near-optimal bounds in three settings.

Result 1 — Junta distributions

Theorem 1. A k-junta distribution $p : \{-1,1\}^n \to [0,1]$ can be learned to d_{TV}-distance ε from

$$O\left(\frac{2^k\, k\, \log(n/\delta)}{\varepsilon^2}\right) \text{ samples.}$$

Result 1 — Junta distributions

Theorem 1. A k-junta distribution $p : \{-1,1\}^n \to [0,1]$ can be learned to d_{TV}-distance ε from

$$O\left(\frac{2^k \, k \, \log(n/\delta)}{\varepsilon^2}\right) \text{ samples.}$$

	Previous (COLT'16)	This work
Upper bound	$O\left(2^{2k} \log n/\varepsilon^4\right)$	$O\left(2^k \log n/\varepsilon^2\right)$
Lower bound	$\Omega\left(2^k/\varepsilon^2 + k \log n/\varepsilon\right)$	matched in every parameter

Result 1 — Junta distributions

Theorem 1. A k-junta distribution $p : \{-1,1\}^n \to [0,1]$ can be learned to d_{TV}-distance ε from

$$O\left(\frac{2^k\, k\, \log(n/\delta)}{\varepsilon^2}\right) \text{ samples.}$$

	Previous (COLT'16)	This work
Upper bound	$O\left(2^{2k} \log n/\varepsilon^4\right)$	$O\left(2^k \log n/\varepsilon^2\right)$
Lower bound	$\Omega\left(2^k/\varepsilon^2 + k \log n/\varepsilon\right)$	matched in every parameter

Quadratic improvement in both 2^k and ε. Algorithm is proper, runs in time $O(n^k 2^k/\varepsilon^2)$.

Result 2 — Quantum junta states (new object)

An n-qubit k-**junta state**:

$$\rho \;=\; \rho_K \otimes \frac{\mathbb{I}_{[n]\setminus K}}{2^{n-k}}, \qquad |K| = k.$$

Nontrivial on k qubits; maximally mixed on the rest. No prior literature.

Result 2 — Quantum junta states (new object)

An n-qubit k-**junta state**:

$$\rho \;=\; \rho_K \otimes \frac{\mathbb{I}_{[n]\setminus K}}{2^{n-k}}, \qquad |K| = k.$$

Nontrivial on k qubits; maximally mixed on the rest. No prior literature.

Theorem 2 (learning). $O\big(12^k \log(n/\delta)/\varepsilon^2\big)$ copies suffice (Pauli measurements, single copies).

Lower bound: $\Omega\big((4^k + \log n)/\varepsilon^2\big)$ copies.

Result 2 — Quantum junta states (new object)

An n-qubit k-**junta state**:

$$\rho \;=\; \rho_K \otimes \frac{\mathbb{I}_{[n]\setminus K}}{2^{n-k}}, \qquad |K| = k.$$

Nontrivial on k qubits; maximally mixed on the rest. No prior literature.

Theorem 2 (learning). $O\!\left(12^k \log(n/\delta)/\varepsilon^2\right)$ copies suffice (Pauli measurements, single copies).

Lower bound: $\Omega\!\left((4^k + \log n)/\varepsilon^2\right)$ copies.

Theorem 3 (testing). For constant k,

$$\widetilde{\Theta}\left(2^n/\varepsilon^2\right) \text{ copies}$$

are necessary and sufficient to test ε-close vs 7ε-far from being a k-junta state.

Result 3 — QAC^0 circuits

Theorem 4 Let ρ be the Choi state of a QAC^0 circuit of size s, depth d, a auxiliary qubits. Then

$$2^{O\left((\log(s^2 2^a/\varepsilon))^d\right)} \log(n/\delta) \text{ copies}$$

suffice to learn ρ in 2^n-Frobenius norm.

Result 3 — QAC^0 circuits

Theorem 4 Let ρ be the Choi state of a QAC^0 circuit of size s, depth d, a auxiliary qubits. Then

$$2^{O\left((\log(s^2 2^a/\varepsilon))^d\right)} \log(n/\delta) \text{ copies}$$

suffice to learn ρ in 2^n-Frobenius norm.

Exponential improvement over Nadimpalli et al. (STOC'24): $n^{O((\log(s/\varepsilon))^d)}$ copies.

Result 3 — QAC^0 circuits

Theorem 4 Let ρ be the Choi state of a QAC^0 circuit of size s, depth d, a auxiliary qubits. Then

$$2^{O\left((\log(s^2 2^a/\varepsilon))^d\right)} \log(n/\delta) \text{ copies}$$

suffice to learn ρ in 2^n-Frobenius norm.

Exponential improvement over Nadimpalli et al. (STOC'24): $n^{O((\log(s/\varepsilon))^d)}$ copies.

In the regime $s = \mathrm{poly}(n)$, $d = O(1)$, $\varepsilon = \Theta(1)$:

- Prior: $n^{\mathrm{polylog}(n)}$ copies.
- Ours: $\log n$ dependence on n (quasi-polynomial overall, but n collapses).

Why the algorithm works

Step 1. Estimate Fourier / Pauli coefficients

$$\widehat{p}(S) \;=\; \mathbb{E}_{x\sim p}\left[\prod_{i\in S} x_i\right], \qquad \widehat{\rho}(P) \;=\; \mathrm{tr}[\rho P]/2^n.$$

Only need degrees $\leq k$ ($\Rightarrow$ low-degree concentration).

Why the algorithm works

Step 1. Estimate Fourier / Pauli coefficients

$$\widehat{p}(S) = \mathbb{E}_{x\sim p}\left[\prod_{i\in S} x_i\right], \qquad \widehat{\rho}(P) = \mathrm{tr}[\rho P]/2^n.$$

Only need degrees $\leq k$ ($\Rightarrow$ low-degree concentration).

Step 2. Threshold small estimates to zero ($\Rightarrow$ sparsity).

Why the algorithm works

Step 1. Estimate Fourier / Pauli coefficients

$$\widehat{p}(S) \;=\; \mathbb{E}_{x\sim p}\left[\prod_{i\in S} x_i\right], \qquad \widehat{\rho}(P) \;=\; \mathrm{tr}[\rho P]/2^n.$$

Only need degrees $\leq k$ ($\Rightarrow$ low-degree concentration).

Step 2. Threshold small estimates to zero ($\Rightarrow$ sparsity).

Step 3. Reconstruct. Parseval + Hoeffding + union bound $\Rightarrow$ error $\leq \varepsilon$.

Why the algorithm works

Step 1. Estimate Fourier / Pauli coefficients

$$\widehat{p}(S) \;=\; \mathbb{E}_{x\sim p}\left[\prod_{i\in S} x_i\right], \qquad \widehat{\rho}(P) \;=\; \mathrm{tr}[\rho P]/2^n.$$

Only need degrees $\leq k$ ($\Rightarrow$ low-degree concentration).

Step 2. Threshold small estimates to zero ($\Rightarrow$ sparsity).

Step 3. Reconstruct. Parseval + Hoeffding + union bound $\Rightarrow$ error $\leq \varepsilon$.

Quantum case: step 1 uses classical shadows.
We give a new proof of optimal shadows via Pauli analysis (non-commutative Bohnenblust–Hille; Volberg–Zhang 2023). No median-of-means needed.

Why QAC^0 enters the story

Theorem 5. The Choi state ρ of an $(n{+}a{+}1)$-qubit QAC^0 circuit of depth d, size s satisfies, for some K of size $|K| \leq (\log(2^a s^2/\varepsilon))^d$,

$$\sum_{\mathrm{supp}(P)\not\subseteq K} |\widehat{\rho}(P)|^2 \;\leq\; \tfrac{\varepsilon}{2^{2n}}.$$

i.e. QAC^0 Choi states are close to *junta* states — stronger than the low-degree concentration of Nadimpalli et al.

Why QAC^0 enters the story

Theorem 5. The Choi state ρ of an $(n+a+1)$-qubit QAC^0 circuit of depth d, size s satisfies, for some K of size $|K| \le (\log(2^a s^2/\varepsilon))^d$,

$$\sum_{\mathrm{supp}(P)\not\subseteq K} |\widehat{\rho}(P)|^2 \;\le\; \tfrac{\varepsilon}{2^{2n}}.$$

i.e. QAC^0 Choi states are close to *junta* states — stronger than the low-degree concentration of Nadimpalli et al.

Theorem 6. A QAC^0 circuit computing the D-address function to error $1/4$ must satisfy

$$s^2\,2^a \;=\; \Omega\!\left(2^{(2^D)^{1/d}}\right).$$

A new lower bound for a canonical low-degree function that is *far from any junta*.

Takeaways

- **One lens, three results:** low-degree $+$ sparse $\Rightarrow$ close to a junta.

Takeaways

- **One lens, three results:** low-degree $+$ sparse $\Rightarrow$ close to a junta.
- **Classical juntas:** optimal $2^k \log n/\varepsilon^2$ samples and quadratic improvement.

Takeaways

- **One lens, three results:** low-degree + sparse $\Rightarrow$ close to a junta.
- **Classical juntas:** optimal $2^k \log n/\varepsilon^2$ samples and quadratic improvement.
- **Quantum junta states:** first study; learning + testing nearly tight.

Takeaways

- **One lens, three results:** low-degree + sparse $\Rightarrow$ close to a junta.
- **Classical juntas:** optimal $2^k \log n/\varepsilon^2$ samples and quadratic improvement.
- **Quantum junta states:** first study; learning + testing nearly tight.
- QAC^0 **circuits:** exponentially fewer copies and new AC^0-style lower bounds.

Takeaways

- **One lens, three results:** low-degree $+$ sparse $\Rightarrow$ close to a junta.
- **Classical juntas:** optimal $2^k \log n/\varepsilon^2$ samples and quadratic improvement.
- **Quantum junta states:** first study; learning $+$ testing nearly tight.
- QAC^0 **circuits:** exponentially fewer copies and new AC^0-style lower bounds.

Takeaways

- **One lens, three results:** low-degree + sparse $\Rightarrow$ close to a junta.
- **Classical juntas:** optimal $2^k \log n/\varepsilon^2$ samples and quadratic improvement.
- **Quantum junta states:** first study; learning + testing nearly tight.
- QAC^0 **circuits:** exponentially fewer copies and new AC^0-style lower bounds.

Thank you!