Random access codes via quantum contextual redundancy

Giancarlo Gatti,[1,2,*] Daniel Huerga,[1,†] Enrique Solano,[1,3,4,5,‡] and Mikel Sanz[1,4,5,§]

[1] *Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, 48080 Bilbao, Spain*
[2] *Quantum Mads, Uribitarte Kalea 6, 48001 Bilbao, Spain*
[3] *International Center of Quantum Artificial Intelligence for Science and Technology (QuArtist) and Department of Physics, Shanghai University, 200444 Shanghai, China*
[4] *IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain*
[5] *IQM, Nymphenburgerstr. 86, 80636 Munich, Germany*

(Dated: September 8, 2021)

We propose a protocol to encode classical bits in the measurement statistics of a set of parity observables, leveraging quantum contextual relations for a random access code task. The intrinsic information redundancy of quantum contexts allows for a posterior decoding protocol that requires few samples when encoding the information in a set of highly entangled states, which can be generated by a discretely-parametrized quantum circuit. Applications of this protocol include algorithms involving storage of large amounts of data but requiring only partial retrieval of the information, as is the case of decision trees. This classical-to-quantum encoding is a compression protocol for more than 18 qubits and shows quantum advantage over state-of-the-art information storage capacity for more than 44 qubits. In particular, systems above 100 qubits would be sufficient to encode a brute force solution for games of chess-like complexity.

Redundancy in classical and quantum information is generally used towards error-correction and data compression. It allows efficient error-correction schemes in order to protect [1, 2] or compress [3] classical data. In quantum systems, different strategies based on classical redundancy [4, 5] or non-local storage of information have been proposed for compressing [6] and error-correcting [7, 8] quantum data.

The fact that not all imaginable outcomes of a measurement context –i.e. a closed set of commuting observables– are possible is summarized by the Kochen-Specker theorem [9] and can be regarded as a source of redundancy susceptible of being used for a quantum compression method of classical data. According to quantum teleportation [10] and superdense coding [11], classical capacity of a quantum channel has at most a 2:1 ratio when storing information digitally. However, statistical approaches may yield better results if only part of the data stored is to be retrieved.

A Random Access Code (RAC) is a communication task where a bitstring is encoded into less information units, and then a bit (or a few bits) of the original message, chosen a posteriori, are retrieved back with a certain success probability. Quantum RACs (QRACs) [12] and entanglement-assisted RACs (EARACs) [13] encode the information into qubits and bits assisted with pre-shared entangled states (ebits), respectively. Current proposals have been shown to achieve a slight quantum advantage for small systems (less than three qubits), but with a fast-decaying success probability as system size increases [14, 15].

In this Article, we propose a QRAC protocol which encodes classical N-bitstrings in the measurement statistics of sets of commuting –i.e. contextual– parity observables (POs). The measurement-basis selection is integrated in the retrieval protocol, taking a middle ground between single-basis and full-tomography approaches. In particular, the quantum register is comprised of n-qubit systems in entangled eigenstates of redundant PO contexts, allowing for an efficient posterior bit retrieval. Selecting a small set of eigenstates with parity-statistics resembling the bitstring, we minimize the sampling requirement (SR) for bit retrieval. Furthermore, we provide with a statistical analysis on the efficiency of this protocol and show that $O(n(3/2)^n)$ n-qubit states can store $O(3^n)$ bits and that any context of bits can be retrieved at a time with $O((3/2)^n)$ samples. For $n \geq 18$, retrieving one context of bits with high fidelity requires less two-level systems than direct data transmission, achieving compression. Moreover, since classical information is encoded in the statistics of POs, the presence of local noise will at most increase the SR, but not corrupt the stored information.

Parity observables.– Let us consider the set of 3^n n-body Pauli observables,

$$\mathcal{O} = \{X, Y, Z\}^{\otimes n}. \tag{1}$$

Each observable within this set, $\mathcal{O}_k \in \mathcal{O}$, with $0 \leq k \leq 3^n - 1$, has two eigenvalues, ± 1, hence both eigenspaces are massively degenerate with dimension 2^{n-1}. In particular, the observable $\mathcal{O}_0 = \otimes_i^n Z_i$ splits its eigenbasis, the so-called computational basis, $|\mathbf{z}\rangle = \otimes_i |z_i\rangle$ with $|z_i\rangle = \{|0\rangle, |1\rangle\}$ [16], in states with even or odd parity. The rest show an equivalent bipartition of their eigenbases. Thus, we will generically call operators in Eq. (1) parity observables.

Measuring the POs on various copies of a generic n-qubit state yields 3^n discrete probability distributions of size two, $\{P_k^+, P_k^-\}$, where $P_k^\pm$ refers to the probability of measuring ± 1 on the kth PO. Considering states with

preferred parities (i.e. either $P_k^+ > 1/2$ or $P_k^- > 1/2$) for all k, there are a total of $2^{(3^n)}$ preferred parity configurations (PPCs) susceptible to storing 3^n bits of information which can be arbitrarily ordered into a parity bitstring.

As a first approach, we may encode a 3^n-bitstring $b = \{b_k\} \mid b_k = \{0,1\}$ in a mixed state of 3^n PO-eigenstates, $\rho = \sum_k p_k |\psi_k\rangle \langle\psi_k|$, each with a well-defined parity matching the PPC, $\mathcal{O}_k |\psi_k\rangle = (2b_k - 1)|\psi_k\rangle$. However, mixed states with exponential number of terms are not efficient as probability distributions rapidly tend to uniform values ($P_k^+ \simeq 1/2$) and thus imply a large SR for its posterior bit retrieval, i.e. the number of copies to sample in order to resolve the preferred parities with a certain success probability, or fidelity [21].

Here we provide instead a more efficient approach leveraging the use of quantum contexts [9] of POs. These are partially joint subsets of commuting POs sharing a common set of eigenstates –i.e. context eigenstates– with well-defined parities in the POs comprising the context and completely undefined in the rest. These states are maximally entangled and allow to encode multiple well-defined parities. We say that a context-PPC is *redundant* if it is representable by a context eigenstate with matching well-defined parities. In general, PPCs with a large number of redundant contexts (RCs) imply a low SR for bit retrieval. To illustrate this, we generate 10^6 instances of 2-qubit random mixed states and classify them according to their PPCs. For each state, we compute the SR of bit k as the number of samples to guarantee with fidelity f_ξ up to rounding precision ξ that the majority yield the preferred parity,

$$S_k(p, f_\xi) = \arg\min_{s\in\mathbb{N}} \left(\sum_{j=\lfloor s/2 \rfloor + 1}^{s} B(s,p;j) > f_\xi \right), \quad (2)$$

where $p = \max(P_k^+, P_k^-)$ and $B(s,p;j) = \binom{s}{j} p^j (1-p)^{s-j}$ is the Binomial density function for j successes in s shots with probability p. We average over all bits, $\overline{S} = \sum_k S_k/3^n$, and select as representative state of each PPC the one with the minimum SR-average, $\overline{S}^\star$.

In Fig. 1 we plot $\overline{S}^\star$ versus the number of occurrences of each PPC. Notably, PPCs cluster according to their number of RCs and whether the RCs span all POs. Equivalently to the Mermin-Peres magic square [9], it can be shown that RCs are limited to either 1, 3 or 5. We can observe that (*i*) PPCs with more RCs require less samples for bit retrieval, and (*ii*) PPCs with more POs spanned by the RCs occupy a larger fraction of the phase space. Based on these observations, we expect a similar behavior for n-qubit systems, and employ context eigenstates for the encoding protocol.

Encoding and retrieval.– First, in order to enhance the efficiency of the protocol, we make a distinction between the parity bitstring, i.e. the ordered PPC, and the *logical* bitstring to be encoded, $\bar{b} = \{\bar{b}_\ell\}$. Specifically, we choose a 2-to-1 mapping between parity and

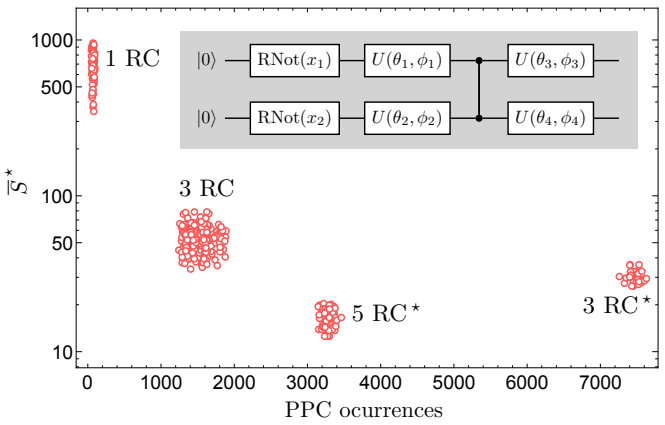

FIG. 1. Lowest SR-average $\overline{S}^\star$ for fidelity $f_\xi = 0.95$ and precision $\xi = 0.005$ versus the number of occurrences of the corresponding PPC in 10^6 random 2-qubit states generated with a parametrized circuit (shown in inset), where $U(\theta,\phi) = R_z(\phi)R_x(-\pi/2)R_z(\theta)R_x(\pi/2)$, and with $\theta \in [0,\pi]$, $\phi \in [0,2\pi]$. Classical randomness allowing the generation of mixed states is introduced through a random not gate: $\mathrm{RNot}(x) = X$ with probability x, and $\mathrm{RNot}(x) = \mathbb{1}$ otherwise. The PPCs are labeled according to their number of RCs and marked with a star when the RCs completely span the whole set of POs.

logical bits by defining an ordered set of PO-couples $\mathcal{G} = \{\{\mathcal{O}_1, \mathcal{O}_2\}, \{\mathcal{O}_3, \mathcal{O}_4\}, ...\}$, leaving the uncoupled PO out. Without loss of generality, we identify the logical bit $\bar{b}_\ell = 0$ (1) with equal (=) and different ($\neq$) preferred parities, respectively, when measuring the corresponding pair of POs. This way, $N = (3^n - 1)/2$ logical bits are encoded in the correlations of pairs of parity bits and we can choose among 2^N PPCs to represent a logical bitstring, which is an exponential degree of freedom. To reduce the retrieval SR, we choose from among all compatible PPCs the one with largest number of RCs or quasi-RCs, i.e. context-PPCs almost matching the statistics of a context eigenstate, as exact representability is increasingly hard with increasing n.

Second, we select a set of context eigenstates with PO eigenvalues resembling the chosen PPC. We consider eigenstates of contexts of largest size, as they are preferable to minimize the retrieval SR, thus serving to upper-bound it and avoid a full characterization of context size and number, which is highly non-trivial with increasing n. The selection of eigenstates can be attained by a discretely-parametrized quantum circuit acting on the Greenberger–Horne–Zeilinger (GHZ) state, $|\mathrm{GHZ}\rangle = (|00\ldots\rangle + |11\ldots\rangle)/\sqrt{2}$ [21]. From a quantum machine learning perspective, this can be understood as a training process in a discrete quantum Born machine [17]. We select the set based on three conditions: *(c1)* possessing at least $\lceil 3^n f^e \rceil$ matches between encoded PPC and target PPC for a given encoding fidelity f^e, *(c2)*

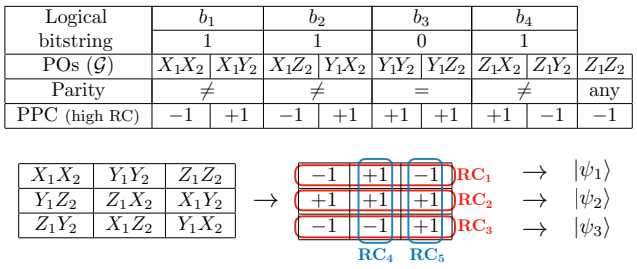

FIG. 2. Example of a 4-bitstring optimally encoded by three 2-qubit states, each one an eigenstate of the row RCs, $|\psi_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|{+}L\rangle - |{-}R\rangle)$ and $|\psi_3\rangle = \frac{1}{\sqrt{2}}(|R{-}\rangle + |L{+}\rangle)$, where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, $|R\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle)$, $|L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle)$. Given an ordered set of PO-couples $\mathcal{G}$, the bitstring is encoded in the parity correlations of pairs of POs. Among all possible PPCs satisfying these correlations, we choose one with a high number of RCs.

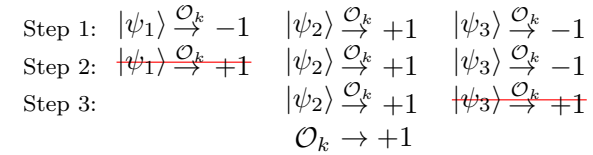

FIG. 3. Retrieval protocol for the preferred parity of $\mathcal{O}_k = Z_1X_2$ from the set of context eigenstates employed in Fig. 2. The protocol consists in filtering out eigenstates with undefined statistics in that PO by performing successive measurements and discarding those that change their outcome.

minimizing the set size N_s, and *(c3)* minimizing the SR-average $\overline{S}$ of the corresponding equiprobabilistic mixed state. These conditions can be summarized in the minimization of a cost function proportional to N_s and to a Normal approximation of the mixed-state SR averaged over the $\lceil 3^n f^\mathrm{e} \rceil$ best-matching POs [21]. We exemplify an optimal encoding of a 4-bitstring in three 2-qubit context eigenstates in Fig. 2.

The retrieval protocol for the kth PO consists in a filtering process where we sequentially measure over the set of encoding states enough times to statistically distinguish well-defined from undefined parities, as shown in Fig. 3. Specifically, we measure the states a given number of times T_k. If a state yields a changing outcome, it is filtered out as it has an undefined parity. The most frequent outcome among unfiltered states defines the preferred parity. Repeating this process for part(all) of the PPC and given the information about the specific order of PO-couples, the bitstring can be partially(completely) retrieved. Notice that the only error source is the residual set of unfiltered undefined-parity states.

Efficiency analysis.– We measure the efficiency of this protocol via a statistical analysis of both the retrieval SR and the fidelity between retrieved and encoded PPCs.

First of all, we need an estimate of the number of eigenstates needed to encode a bitstring, N_s. For that purpose, we model the number of matches between an arbitrary context eigenstate and an arbitrary context-PPC by a binomial distribution, $x \sim B(d_\mathcal{C}, 1/2; \cdot)$, where $d_\mathcal{C}$ refers to the context size, and we consider that each well-defined parity of the state matches the corresponding preferred parity with probability 1/2. Thus, a context eigenstate has probability $p_h = \sum_{k=0}^{h} B(d_\mathcal{C}, 1/2; k)$ of matching the PPC in the POs of its context with up to h mismatches (i.e. so-called Hamming distance), with a lower bound $p'_h = \binom{d_\mathcal{C}}{h} 2^{-d_\mathcal{C}}$. For even number of qubits $n > 4$, there are $N_\mathcal{C} = 2 \times 3^n$ contexts of maximum size, $d_\mathcal{C} = 2^{n-1} + 1$ [21], each characterized by a basis of 2^n context eigenstates (the size of the Hilbert space). Considering this, we model the total number of states with up to h mismatches in their respective contexts by a binomial distribution, $\mathcal{N}_h \sim B(2^n N_\mathcal{C}, p'_h; \cdot)$. Given a logical bitstring, each of the 2^N compatible PPCs is an instance of $\mathcal{N}_h$, among which we select the most redundant PPC. Therefore, we take the largest N_s such that at least one compatible PPC is expected to have those many eigenstates with up to h mismatches, as given by setting $2^N B(2^n N_\mathcal{C}, p'_h; N_\mathrm{s}) = 1$. Applying Stirling's approximation and keeping the dominant term in n yields

$$N_\mathrm{s}(n, \varepsilon) = \left(\frac{3}{2}\right)^n g(\varepsilon)^{-1}, \tag{3}$$

where we have defined the encoding match error fraction, $\varepsilon = h/d_\mathcal{C}$, and $g(\varepsilon) = 1 - \frac{1-\varepsilon}{\ln(2)} \sum_{i=1}^{\infty} \frac{\varepsilon^i}{i} - \varepsilon \log_2(1/\varepsilon)$. Notice that $N_\mathrm{s} \to (3/2)^n$ when $\varepsilon \to 0$.

Considering $N_k < N_\mathrm{s}$ states with well-defined kth PO, the encoding fidelity is the probability that the majority of them yield the target preferred parity, plus 1/2 the probability that exactly half of them yield it,

$$f_k^\mathrm{e} = \sum_{i=\lfloor N_k/2 \rfloor + 1}^{N_k} B(N_k, 1-\varepsilon;\, i) + \frac{1}{2} B\left(N_k, 1-\varepsilon; \frac{N_k}{2}\right), \tag{4}$$

where $B(N_k, 1-\varepsilon; \frac{N_k}{2}) = 0$ for odd N_k and we take $f_k^\mathrm{e} = 1/2$ for $N_k = 0$. Remark that $f_k^\mathrm{e}(N_k) = f_k^\mathrm{e}(N_k + 1)$ for odd N_k, such that the minimal N_k for any fixed fidelity is always odd.

Regarding the SR, we perform T_k measurement steps in order to retrieve the parity bit of the kth PO and thus a total of $(N_\mathrm{s} - N_k)(1 + \sum_{i=0}^{T_k - 2} 2^{-i}) \approx 3(N_\mathrm{s} - N_k)$ measurements are performed on undefined-parity states. In the process, an average of $\nu_k = (N_\mathrm{s} - N_k)/2^{T_k - 1}$ states fail to be filtered out, which can be interpreted as the probability of having a single unfiltered state if $\nu_k \ll 1$, i.e. as the retrieval noise probability. By inverting this relation, $T_k(n_k, \nu_k) = \log_2(N_\mathrm{s} - N_k) - \log_2(\nu_k) + 1$ sets an upper bound to the number of measurements performed on any single state. Therefore, the SR to retrieve a parity configuration on the kth PO by measuring over the whole

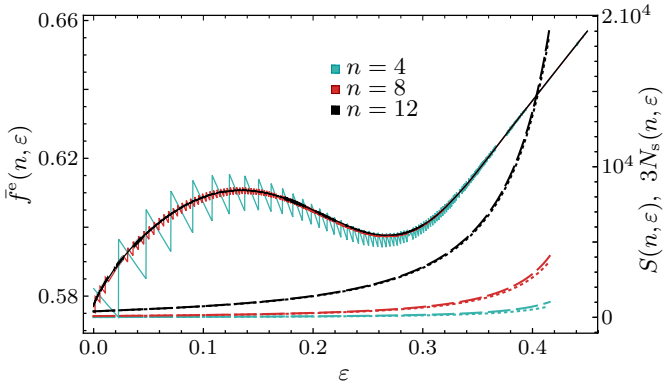

FIG. 4. Left axis: Expected logical bit encoding fidelity (solid line) for $n = 4, 8, 12$ (cyan, red, and black, respectively). Right axis: expected SR (dashed line) for $\nu_k = \nu = 2^{-6} \ll (1 - f^{\rm e})$ and the eigenstate selection size, $N_{\rm s}$, multipied by 3 (dotted line), for same system sizes. Notice that S and $3N_{\rm s}$ are approximately equal, implying that in average 3 measurements are required per context eigenstate (see Eq. 5).

set of states is approximately

$$S_k(n, N_k, \varepsilon, \nu_k) = 3\left[N_{\rm s}(n,\varepsilon) - N_k\right] + T_k N_k. \qquad (5)$$

Notice that the retrieval noise ν_k can be set to a negligible value (with respect to encoding errors) without increasing S_k too much, so that the overall fidelity depends only on encoding errors.

The expected parity bit encoding fidelity and the SR are obtained from Eqs. (4) and (5), respectively, by modeling $N_k \sim B(\lfloor N_{\rm s} \rfloor, p'_k; \cdot)$, where $p'_k = (1/2)(2/3)^n$ approximates the probability that a context eigenstate contains the kth PO in its context, $p'_k < p_k = d_{\mathcal{C}}/3^n$. Considering that a logical bit is retrieved successfully when both of its associated parities are correct or both are wrong, the logical encoding fidelity is $\bar{f}^{\rm e} = (f^{\rm e})^2 + (1 - f^{\rm e})^2$ for a typical parity encoding fidelity.

In Fig. 4, we plot the logical encoding fidelity and expected SR with respect to the encoding error for various system sizes ($n = 4, 8, 12$), setting a negligible $\nu_k = \nu \ll (1 - f^{\rm e})$, which allows to approximate the overall fidelity by the encoding fidelity alone ($\bar{f} \approx \bar{f}^{\rm e}$). Discontinuities in the fidelity tend to vanish for large number of selected states (3), such that the fidelity converges into a single curve for large number of qubits ($n \gtrsim 8$) and large encoding errors ($\varepsilon \gtrsim 0.3$), as they both increase the number of selected eigenstates. When increasing the error tolerance, the increase of the number of selected eigenstates favors higher fidelity but at the same time hinders it because their associated parity bitstrings have a poorer resemblance to the target PPC. This dual behaviour explains the local minimum of the fidelity.

Notice that the fidelity local maximum is not necessarily the optimal value for the encoding error tolerance, as the SR must be taken into account too. In order to obtain the encoding error fraction that minimizes the SR for a given fixed fidelity, we perform r repetitions of the encoding-and-retrieval protocol such that the fidelity is fixed for all values of ε (each one requires a different value of r), and then minimize the SR. In this way, we obtain an optimal error tolerance $\varepsilon^\star = 0.0480$ for large n. In general, logical bit fidelity $\bar{f}_r = \sum_{j=(r+1)/2}^{r} B(r, \bar{f}; j)$ is reached with odd r repetitions, and in particular $r = 239$ suffices to achieve $\bar{f}_r = 99.9\%$ for $n \geq 16$. Furthermore r can be isolated in terms of the fidelity by taking the Normal approximation [21].

The optimal number of measurement steps can be written as a function of n and upper bounded by

$$T^\star(n) = n \log_2(3/2) + 9, \qquad (6)$$

where we have considered $N_k \sim O(N_{\rm s} p'_k) \sim O(1/2)$ and $\nu_k = \nu \ll (1 - f^{\rm e})$ with $2^{-7} \leq \nu \leq 2^{-5}$. Recall that Eq. (6) indicates the maximum number of measurements performed on any state of the set and thus is the number of copies we need of each context eigenstate in order to correctly retrieve any parity bit. To retrieve the parity bits of a group of POs, we apply $T^\star$ steps of the retrieval protocol on each PO. If two POs commute, any state measured on one of them can still be measured on the other. Thus, any context of POs can be measured with $T^\star$ copies of the set of states. In order to reduce the number of copies required to retrieve the logical bits, it is convenient to choose PO orderings where most of the PO couples commute.

Equivalently, we can express the expected SR (5) as a function of the system size,

$$S^\star(n) = g(\varepsilon^\star)^{-1}\left[3\left(\frac{3}{2}\right)^n + \frac{T^\star(n)}{2} - \frac{3}{2}\right]. \qquad (7)$$

Discussion and conclusion.– The present protocol uses a total of $Q(n, r) = r\, n\, N_{\rm s}(n, \varepsilon^\star)\, T^\star(n)$ two-level systems separable in n-qubit batches to encode $N = (3^n - 1)/2$ logical bits and retrieve one context of preferred parities, that is, $d_{\mathcal{C}} = 2^{n-1} + 1$ parity bits. While directly reading any portion of the data requires the availability of as many two-level systems as the number of logical bits, i.e. $N \sim O(3^n)$, the present protocol only needs $Q \sim O(n^2 (3/2)^n \ln(1/(2 - 2\bar{f})))$ qubits. Specifically, considering $r = 239$ repetitions, which implies 99.9% fidelity for $n \geq 16$, the protocol uses less two-level resources than direct transmission for $n \geq 18$. In this case, it can be considered a data compression protocol, i.e. $Q < N$, showing large-scale quantum advantage with non-decaying fidelity, in contrast to previous QRAC proposals [14, 15]. Up to $O(2^{n-1}/n^2)$ contexts of preferred parities can be retrieved and maintain this advantage.

Remarkably, the degree of compression achieved is large enough to store a whole cloud-storage server of 1 billion users [18] with $\sim$ 800 Gb each in systems of 44 qubits. Specifically, to retrieve one context of parity bits with 99.9% fidelity, we need $\sim 6 \times 10^{11}$ of such

44-qubit systems, i.e. 26.4 kilo-qubits per user. Furthermore, the protocol is readily applicable for algorithms involving large decision trees, e.g. brute-force strategies, as only one branch needs to be consulted at a time. A brute-force solution for problems of chess-like complexity would require 100-qubit systems to store one in $35^{2^{155}}$ possible strategies (2.34×10^{47} bits), considering an upper bound of 2^{155} chess board configurations [19] and a branching factor of ~ 35 possible plays per turn [20].

Acknowledgements.– The authors acknowledge financial support from Spanish Government PGC2018-095113-B-I00 (MCIU/AEI/FEDER, UE), Basque Government IT986-16, as well as from QMiCS (820505) and OpenSuperQ (820363) of the EU Flagship on Quantum Technologies, EU FET Open Grant Quromorphic (828826), EPIQUS (899368) and Shanghai STCSM (Grant No. 2019SHZDZX01-ZX04)

* gatti.gianc@gmail.com
† huerga.daniel@gmail.com
‡ enr.solano@gmail.com
§ mikel.sanz@ehu.eus

[1] C. E. Shannon, *A mathematical theory of communication*, The Bell system technical journal **27**, 379–423 (1948).
[2] W. C. Huffman and V. Pless, *Fundamentals of error-correcting codes* (Cambridge University Press, 2012).
[3] H. Al-Bahadili, *A novel lossless data compression scheme based on the error correcting Hamming codes*, Computers & Mathematics with Applications **56**, 143–150 (2008).
[4] A. R. Calderbank and P. W. Shor, *Good quantum error-correcting codes exist*, Phys. Rev. A **54**, 1098–1105 (1996).
[5] A. M. Steane, *Error correcting codes in quantum theory*, Phys. Rev. Lett. **77**, 793–797 (1996).
[6] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, *Quantum data compression of a qubit ensemble*, Phys. Rev. Lett. **113**, 160504 (2014).
[7] D. Gottesman, *Class of quantum error-correcting codes saturating the quantum Hamming bound*, Phys. Rev. A **54**, 1862–1868 (1996).
[8] A. Y. Kitaev, *Fault-tolerant quantum computation by anyons*, Annals of Physics **303**, 2–30 (2003).
[9] A. Peres, *Quantum theory: Concepts and Methods* (Springer Science & Business Media, 2006).
[10] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, Phys. Rev. Lett. **70**, 1895 (1993).
[11] C. H. Bennett and S. J. Wiesner, *Communication via one-and two-particle operators on Einstein-Podolsky-Rosen states*, Phys. Rev. Lett. **69**, 2881 (1992).
[12] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, *Dense quantum coding and a lower bound for 1-way quantum automata*, in Proceedings of the thirty-first annual ACM symposium on Theory of Computing (1999) pp. 376–383.
[13] M. Pawłowski and M. Żukowski, *Entanglement-assisted random access codes*, Phys. Rev. A **81**, 042326 (2010).
[14] A. Casaccino, E. F. Galvão, and S. Severini, *Extrema of discrete Wigner functions and applications*, Phys. Rev. A **78**, 022310 (2008).
[15] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, *Quantum random access codes using single d-level systems*, Phys. Rev. Lett. **114**, 170502 (2015).
[16] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
[17] S. Cheng, J. Chen, and L. Wang, *Information perspective to probabilistic modeling: Boltzmann machines versus Born machines*, Entropy **20**, 583 (2018).
[18] F. Lardinois, *Google drive will hit a billion users this week*, TechCrunch (2018).
[19] J. Tromp, *John's chess playground*, (2010).
[20] A. Levinovitz, *The mystery of Go, the ancient game that computers still can't win*, Wired Business (2014).
[21] See Supplementary Material.

Supplementary Material

Counting contexts of largest size

For $n = 2$ and $n = 4$, contexts are constant in size. In particular, for $n = 2$ the 6 contexts are composed of 3 POs each. The POs can be arranged in a 3×3 square where each context corresponds to either a row or a column,

X_1X_2	Y_1Y_2	Z_1Z_2
Y_1Z_2	Z_1X_2	X_1Y_2
Z_1Y_2	X_1Z_2	Y_1X_2

(8)

where $\{X_j, Y_j, Z_j\}$ refer to the three Pauli matrices on qubit j,

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (9)$$

Therefore, every PO belongs to 2 different contexts.

For $n = 4$, the 270 contexts are composed of 9 POs each, but they are not as easily represented as $n = 2$.

For $n > 4$, contexts have different sizes. Largest context sizes allow for well-defined parities in more POs simultaneously, whereas smaller contexts are more numerous and thus allow to select the encoding context eigenstates from a larger pool. For example, for 6-qubit systems, there are 1458 contexts of size 33, 17820 contexts of size 27, and 19440 contexts of size 24. A full characterization of context number and size for n-qubits is non-trivial, thus the encoding protocol that we provide builds upon the use of largest contexts. For that purpose we propose the following algorithm to generate contexts of largest size for even n.

First, choose an arbitrary PO, which we will refer to as a *context generator* (CG). For simplicity, we choose here $\mathcal{O}_0 = \prod_i Z_i$. Any PO sharing no Pauli operator with it in the same qubit positions,

$$\mathcal{C}' = \{X, Y\}^{\otimes n}, \quad (10)$$

commutes with the CG. All POs in $\mathcal{C}'$ do not necessarily commute among themselves. For the specific 2-qubit case, they can be classified into two groups $\{X_1X_2, Y_1Y_2\}$ and $\{X_1Y_2, Y_1X_2\}$, and the two contexts correspond to the first row and last column of Eq. (8). In general, the 2^n POs in $\mathcal{C}'$ can at the same time be classified into two different groups, where the POs of each group commute with each other. This way, we build contexts of largest size $d_{\mathcal{C}} = 2^{n-1} + 1$ (the $+1$ being the generator), two of them for every generator.

Observables of the type $\{X, Y\}^{\otimes n-m} Z^{\otimes m}$ with even $n - m$ also commute with the CG, but share terms with it in the same qubit positions. This case will be addressed later, as it generates a second kind of contexts.

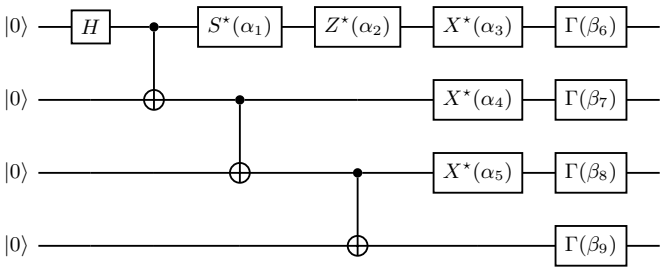

FIG. 5. Discretely-parametrized quantum circuit to generate 4-qubit eigenstates of contexts of maximum size. Here H is the Hadamard gate, and $\alpha = \{0, 1\}$ and $\beta = \{0, 1, 2\}$ are discrete parameters. The gates are defined such that $S^\star(0) = Z^\star(0) = X^\star(0) = \Gamma(0) = \mathbb{1}$ and $S^\star(1) = S = |0\rangle\langle 0| + i\,|1\rangle\langle 1|$, $Z^\star(1) = Z$, $X^\star(1) = X$, $\Gamma(1) = H$ and $\Gamma(2) = S.Z.H$. A similar circuit layout can be used to generate n-qubit context eigenstates (of contexts of largest size) for arbitrary n.

For $n = 2$, the generator is not unique to a context. However, for $n \geq 4$ every context has a unique generator. As an example, let us consider the context

$$\begin{aligned} &\{\hat{Z}_1\,\hat{Z}_2\,\hat{Z}_3\,\hat{Z}_4\,, \hat{X}_1\,\hat{X}_2\,\hat{X}_3\,\hat{X}_4\,, \hat{Y}_1\,\hat{Y}_2\,\hat{X}_3\,\hat{X}_4\,, \\ &\hat{Y}_1\,\hat{X}_2\,\hat{Y}_3\,\hat{X}_4\,, \hat{Y}_1\,\hat{X}_2\,\hat{X}_3\,\hat{Y}_4\,, \hat{X}_1\,\hat{Y}_2\,\hat{Y}_3\,\hat{X}_4\,, \\ &\hat{X}_1\,\hat{Y}_2\,\hat{X}_3\,\hat{Y}_4\,, \hat{X}_1\,\hat{X}_2\,\hat{Y}_3\,\hat{Y}_4\,, \hat{Y}_1\,\hat{Y}_2\,\hat{Y}_3\,\hat{Y}_4\}, \end{aligned} \quad (11)$$

where the CG is the only PO sharing no term with the rest in any of the qubit positions. There are uniquely 2 contexts to which a CG belongs, and thus there is a total $N_{\mathcal{C}} = 2 \times 3^n$ contexts of size $d_{\mathcal{C}} = 2^{n-1} + 1$. Full numerical characterization of context number and size up to $n = 20$ qubits shows that this algorithm produces the contexts of largest size and counts them correctly.

For purposes of the data compression protocol presented in this article, tallying the contexts of maximum size is sufficient, as context number has a much lower contribution than context size.

Since context sizes grow exponentially in n, the total number of possible PPCs grows superexponentially. Specifically, there are $2^{d_{\mathcal{C}}}$ possible PPCs for the contexts of maximum size. This is much greater than the Hilbert space (2^n), which explains why not all context PPCs are representable by context eigenstates.

Generation of context eigenstates

To construct a parametrized quantum circuit that can generate context eigenstates (of contexts of maximum size), we first establish a set of discretely-parametrized gates that can produce the context eigenstates associated to one generator (as defined in the context-counting algorithm above). Without loss of generality, consider the generator $\mathcal{O}_0 = \prod_i \hat{Z}_i$ for n-qubit systems. This generator is associated to two different contexts, and the GHZ state can be shown to be eigenstate to one of them,

whereas the GHZ state with a phase of i (GHZ-i) can be shown to be eigenstate to the other. Pauli rotations X and Z applied on the GHZ (GHZ-i) state produce the eigenstate basis for the corresponding context of that generator.

We can now generate all context eigenstates from one generator. To extend the parametrized circuit to context eigenstates of other generators, we consider a set of transformations to map one Pauli basis to another. This can be achieved by the set of 3^n transformations given by $\{\mathbb{1}, H, S.Z.H\}^{\otimes n}$. Note that the Z basis elements $\{|0\rangle, |1\rangle\}$ are transformed by the Hadamard gate H into the X basis elements $\{|+\rangle, |-\rangle\}$, and by $S.Z.H$ into the Y basis elements $\{|R\rangle, |L\rangle\}$, where S is the i-phase gate.

This way, if we append the set of Pauli-basis transformations into a circuit that can already produce all the context eigenstates of one generator, we will be able to produce all context eigenstates of maximum size. We exemplify this circuit for $n = 4$ in Fig. 5. The circuit is trivially scalable to systems of any size, and is defined in such a way that there are exactly $N_{\mathcal{C}}\, 2^n = 2 \times 6^n$ possible outputs, one for each eigenstate.

Computation of mixed state sampling requirement

Given a state with parity distribution in the form $\{P^+, P^-\}$, it is easier to distinguish the preferred parity of the state if its values are more extreme. E.g., it is easier to learn that $P^+ > P^-$ from a state with the $\{0.8, 0.2\}$ distribution than it is from one with the $\{0.55, 0.45\}$ distribution. Consequently, if a PPC bitstring can be generated with a mixed state of fewer terms (or terms with well-defined parities in more POs), the parity distributions will be more extreme and require fewer samples to be distinguished. To illustrate this in n-qubit systems, let us consider two cases: encoding an arbitrary PPC (size 3^n, not specifically selected to have high number of RCs) in 1) a balanced mixed state of PO-eigenstates and 2) a balanced mixed state of context eigenstates.

PO-eigenstates.– For the first case we make use of a balanced mixed state $\rho = \sum_k \frac{1}{3^n} |\psi_k\rangle \langle\psi_k|$ of 3^n PO-eigenstate terms $|\psi_k\rangle$, with well defined parity in the kth PO and undefined in the rest. This way, the parity distribution of the mixed state for any PO is the average between $(3^n - 1)$ undefined parity distributions and one well-defined distribution, i.e.

$$\{P^+, P^-\} = \left\{ \frac{1}{2}\left(1 \pm \frac{1}{3^n}\right), \frac{1}{2}\left(1 \mp \frac{1}{3^n}\right) \right\}. \tag{12}$$

The corresponding SR S can be computed with Eq. (2) for $p = \max(P^+, P^-)$ and retrieval fidelity $f_{\text{mix}} = 0.84135$ (chosen for reasons explained below) up to rounding error $\xi = 5 \times 10^{-6}$, as shown in Table I. The parity distribution is equal to an undefined distribution save for an exponentially small term, thus the SR to identify the preferred parity is bound to be large. However, actually calculating it is computationally hard with increasing n.

n	S_{PO}	$\tilde{S}_{\text{PO}}$	$\tilde{S}_{\mathcal{C}}$
2	81	80	53.0
3	729	728	72.2
4	6561	6560	181.9
5	59049	59048	554.7
6	-	531440	1856.9
7	-	4.783×10^6	6583.3
8	-	4.305×10^7	24284.5

TABLE I. Sampling requirement exact value S_{PO} and its Normal approximation $\tilde{S}_{\text{PO}}$ to retrieve a preferred parity bit from a mixed state of 3^n PO-eigenstates of n qubits, for fidelity $f_{\text{mix}}(n_\sigma = 1) = 0.84135$. The exact value is calculated only up to $n = 5$ for computational reasons. Sampling requirement Normal approximation $\tilde{S}_{\mathcal{C}}$ to retrieve a preferred parity bit from a mixed state of context eigenstates of n qubits, also for fidelity $f_{\text{mix}}(n_\sigma = 1) = 0.84135$.

To provide a general expression for n qubits, we consider the sum term in Eq. (2), which corresponds to the fidelity of identifying the preferred parity correctly for a given number of samples s. Then, for large s, which is associated to a large enough fidelity f_{mix}, we approximate the binomial distribution $B(s, p; j)$ by a Normal probability density function $N(\mu = sp, \sigma = \sqrt{sp(1-p)})$ and the sum for an integration, such that

$$f_{\text{mix}} \approx \int_{s/2}^{\infty} N(\mu, \sigma). \tag{13}$$

The probability for a sample from a Normal distribution to deviate at most n_σ standard deviations (SD) from the mean in any direction is $f_{\text{SD}}(n_\sigma) = \text{Erf}(n_\sigma/\sqrt{2})$, where $\text{Erf}(x)$ is the error function. Since we deviate $n_\sigma = \frac{\mu - s/2}{\sigma}$ SDs from the mean in the lower integration limit, the fidelity, given by this integral, is $f_{\text{mix}}(n_\sigma) = 1/2 + f_{\text{SD}}(n_\sigma)/2$. Notoriously, $f_{\text{SD}}(1) = 0.6827$, $f_{\text{SD}}(2) = 0.9545$, and $f_{\text{SD}}(3) = 0.9973$. Hence, $f_{\text{mix}}(1) = 0.84135$, which why we chose this fidelity value. We denote the SR Normal approximation by $\tilde{S}$.

This way, the number of samples to retrieve the preferred parity with fidelity $f_{\text{mix}}(n_\sigma)$ is

$$\tilde{S} = n_\sigma^2 \frac{p(1-p)}{(p - 1/2)^2}, \tag{14}$$

and yields for the present case, that is $p = (1 + 1/3^n)/2$,

$$\tilde{S}_{\text{PO}} = n_\sigma^2 (3^{2n} - 1), \tag{15}$$

which is of order $\sim O(9^n)$.

As can be noted from Table I, $\tilde{S}_{\text{PO}}$ is a good approximation of S_{PO} for any n. The value of S is always odd due

to the rounding term in Eq. (2), thus if $\tilde{S}_{\mathrm{PO}}$ is rounded up to the nearest odd value, S_{PO} and $\tilde{S}_{\mathrm{PO}}$ are exactly equal between $n = 2$ and $n = 5$.

Context eigenstates.– For the second case, we consider an arbitrary PPC to be encoded using context eigenstates from $N_{\mathcal{C}} = 2 \times 3^n$ contexts of size $d_{\mathcal{C}} = 2^{n-1} + 1$, corresponding to the contexts of maximum size. Each context has 2^n eigenstates ($N_{\mathcal{C}}\, 2^n$ in total), and every eigenstate partially *matches* the desired PPC in some POs of that context. For an arbitrary eigenstate, each of its well-defined parities has probability $1/2$ of matching the respective bit, thus the total number of matches for a context eigenstate can be modelled by a binomial distribution $B(d_{\mathcal{C}}, 1/2; \cdot)$. Evidently, some context eigenstates will have more matches than others, so we build a mixed state with the best N_{s} eigenstates from all contexts. To do this, we take the $m = N_{\mathrm{s}}/(N_{\mathcal{C}}\, 2^n)$ percentile of them with the greatest number of matches, and estimate a lower bound for the number of matches of this selection, employing the limiting value for that percentile in the Normal distribution approximation. A mixed state of N_{s} context eigenstates with this number of matches yields a typical parity distribution

$$\{P_t^+, P_t^-\} = \left\{ \frac{1}{2}\left(1 \pm \frac{\Delta}{3^n}\right), \frac{1}{2}\left(1 \mp \frac{\Delta}{3^n}\right) \right\}, \tag{16}$$

where $\Delta = \sqrt{2}\,\mathrm{Erfc}^{-1}(2m)\sqrt{d_{\mathcal{C}}}$ is the difference between number of matches and number of mismatches for the lower bound. This difference is proportional to the standard deviation of the Normal distribution, and has a $\mathrm{Erfc}^{-1}(x)$ term (inverse of the complementary error function) from an integration of the Normal distribution in the chosen percentile. The size of N_{s} must be such that all POs have at least one well-defined parity in the collection of context eigenstates, meaning N_{s} should at least satisfy $N_{\mathrm{s}} > 3^n/d_{\mathcal{C}} \sim O((3/2)^n)$. For a finer approximation, we compute the probability of a PO not having any well-defined parity in a set of N_{s} context eigenstates, $\left((3^n - d_{\mathcal{C}})/3^n\right)^{N_{\mathrm{s}}}$, and set it to be much smaller than the retrieval error $(1 - f_{\mathrm{mix}})$, so that it can be ignored. This way, we choose $\left((3^n - d_{\mathcal{C}})/3^n\right)^{N_{\mathrm{s}}} \stackrel{!}{=} e^{-7} \ll (1 - f_{\mathrm{mix}})$. Taking the natural logarithm on both sides, expanding the left logarithm and approximating $\sum_{k=1}^{\infty}(d_{\mathcal{C}}/3^n)^k/k \approx d_{\mathcal{C}}/3^n$ for large n, we obtain $N_{\mathrm{s}} \approx 7 \times 3^n/d_{\mathcal{C}}$. It follows that $m = 7/(2^n(2^n+2))$ and, considering that $\mathrm{Erfc}^{-1}(y) \approx \sqrt{\ln(1/y)}$ for $y \ll 1$, we compute $\mathrm{Erfc}^{-1}(2m) \approx \sqrt{\ln(2^n) + \ln(2^n+2) - \ln(14)} \approx \sqrt{2\ln 2}\sqrt{n}$. Finally, this leads us to $\Delta \approx 2\sqrt{\ln(2)}\sqrt{n d_{\mathcal{C}}}$. Then, applying Eq. (14), the number of samples to retrieve the preferred parity with fidelity $f_{\mathrm{mix}}(n_\sigma)$ is approximately

$$\tilde{S}_{\mathcal{C}} = \frac{n_\sigma^2}{2}\left(\frac{9^n}{2^n n \ln(2) + n\ln(4)} - 2\right) \sim O\left(\frac{(9/2)^n}{n}\right). \tag{17}$$

For $n_\sigma = 1$ the fidelity is $f_{\mathrm{mix}}(n_\sigma) = 0.84135$, in which case $\tilde{S}_{\mathcal{C}}$ takes the values shown on Table I. As can be noted, the SR to retrieve a PPC encoded with a mixed state of context eigenstates is orders of magnitude lower than when using a mixed state of PO-eigenstates, $O((9/2)^n/n)$ versus $O(9^n)$, respectively.

Context eigenstates are maximally entangled states (specifically GHZ states for the contexts of largest size) whereas PO-eigenstates can always be unentangled states. Thus, the reduction in the SR is due to taking advantage of quantum resources. However, neither of these approaches has better SR than directly measuring the bitstring. Namely, if a bistring of size 3^n is encoded in 3^n two-level systems, someone wanting to retrieve an arbitrary bit would need to have all 3^n states available for measurement. A single measurement would be required, but it would be from choosing one in a collection of 3^n states. Thus, the encoding protocol can not be considered a *compression* until this problem is addressed, even if the SR is reduced by means of quantum resources.

The use of mixed states is the reason why neither of these encodings has better SR than the number of states required for a straightforward retrieval. Because of this, we opt to use *sets of states* and a retrieval protocol instead of mixed states and blind sampling, since, as shown in the main text, the former reaches a SR of $O(n(3/2)^n)$, which allows to achieve compression for sufficiently large n.

Encoding Implementation

In the following, we show cost functions that can be used to select a high-RC PPC for a given logical bitstring and afterwards build a selection of context eigenstates to encode it. Without loss of generality, we apply these cost functions to encode a logical bitstring of $N = 40$ binary digits of $\pi/4$ with eigenstates of $n = 4$ qubits from contexts of maximum size ($d_{\mathcal{C}} = 9$), choosing alphabetical $\mathcal{G} = \{\{X_1X_2X_3X_4, X_1X_2X_3Y_4\}, ...\}$ ordering.

PPC selection.– We first compute the well-defined parities of all eigenstates and select a high-RC PPC compatible with the logical bitstring. We do this maximizing a scoring function $\mathcal{L} = \sum_i w_i \mathcal{M}_i$ based on the number $\mathcal{M}_i$ of context eigenstates with i matches, with weights $\{w_i\}$. To prioritize maximizing the number of full-match RCs and then select a compatible PPC with a high number of contexts approaching a full-match, we choose a high weight value for full matches, significantly lower values for the number of states missing a few matches, and weight zero for states approaching half the matches or

less. Then, we discretely parametrize the set of compatible PPCs with a bitstring and determine a local maxima for the scoring function.

Specifically for our example, we chose $w_9 = 100$, $w_8 = 10$, $w_7 = 1$, and $w_i = 0$ for $i \leq 6$, obtaining a compatible PPC with $\mathcal{M}_9 = 30$, $\mathcal{M}_8 = 51$ and $\mathcal{M}_7 = 78$ from the local maximization.

Algorithm 1: State Selection

Set number of training steps T (index t)
Define bit-vector $\mathbf{v}$ of size κ_s
Set number of training steps $n_{\mathscr{L}}$ and number of random bit-flips n_r that will alternate
Define cost function $\mathscr{L}(\mathbf{v})$ (Eq. (19))
Define a function to flip individual bits of $\mathbf{v}$:
$\mathbf{flip}_k(\mathbf{v}) = \{v_1, v_2, ..., v_{k-1}, 1 - v_k, v_{k+1}, ..., v_{\kappa_s}\}$

$\mathbf{v}^{(0)} \leftarrow \{1, ..., 1\}$
$t \leftarrow 0$
$t' \leftarrow 0$
while $t < T$ **do**
  if $t' > 0$ **then**
    for all $1 \leq j \leq n_r$ **do**
      $r \leftarrow$ **Random**$[\{1, 2, ..., N_s\}]$
      $\mathbf{v}^{(t+1)} \leftarrow \mathbf{flip}_r(\mathbf{v}^{(t)})$
      $t \leftarrow t + 1$
    end for
  end if
  for all $1 \leq j \leq n_{\mathscr{L}}$ **do**
    Find $k \in [1, \kappa_s]$ minimizing $\mathscr{L}(\mathbf{flip}_k(\mathbf{v}^{(t)}))$ and assign it to k'
    $\mathbf{v}^{(t+1)} \leftarrow \mathbf{flip}_{k'}(\mathbf{v}^{(t)})$
    $t \leftarrow t + 1$
  end for
  $\mathbf{v}'^{(t'+1)} \leftarrow \mathbf{v}^{(t)}$
  $t' \leftarrow t' + 1$
end while
$T' \leftarrow t'$
Find $i \in [1, T']$ minimizing $\mathscr{L}(\mathbf{v}'^{(i)})$ and assign it to i'
Return $\mathbf{v}'^{(i')}$
end

Eigenstate selection.– Afterwards, we select a small set of eigenstates that encodes the PPC $\{b_k\}$ with various degrees of fidelity f^e (i.e. $\lceil 3^n f^e \rceil$ POs encoded with the correct parity) and minimizes the retrieval SR. To do this, we define a bit-vector to indicate which states are part of the set, $\mathbf{v} = \{v_i\}$ with $v_i = \{0, 1\}$ for $i = \{1, ..., \kappa_s\}$, where κ_s is the size of the pool of eigenstates we are selecting from, discarding beforehand those with less than half well-defined parities matching the PPC. If $v_i = 1$, it means that context eigenstate $|\psi_i\rangle$ is part of the set. This way, the size of the selected set is $N_s = \sum_i v_i$. To perform the selection, we define $\mathbf{v}^{(t)}$ as the value of $\mathbf{v}$ at step t of the algorithm. Starting with $\mathbf{v}^{(0)} = \{1, ..., 1\}$, we update $\mathbf{v}$ at each step to minimize a cost function related to the average-SR of the corresponding mixed state $\rho(\mathbf{v}) = \sum_{i=1}^{N_s} v_i |\psi_i\rangle \langle\psi_i| / N_s$. For this, we approximate the SR with a variation of Eq. (14),

$$S'(p) = \begin{cases} \text{Min}(u,\, p(1-p)/(p-1/2)^2) & p > 1/2 \\ u & p \leq 1/2 \end{cases}, \quad (18)$$

for an arbitrarily high u to prevent divergence of $S'(p)$, and choosing $n_\sigma = 1$ without loss of generality, as it is a term that affects all values of $\mathbf{v}$ equally. Note that Eq. (14) implicitly assumed $p \geq 1/2$ when taking $p = Max(P^+, P^-)$, but this is no longer the case. Instead, we use $p_k(\mathbf{v}) = |1 - b_k - P_k^+(\rho(\mathbf{v}))|$ such that $p_k \leq 1/2$ when there is a mismatch between preferred parity and target parity, in which case we set $S'(p_k) = u$. Recall that $P_k^+(\rho)$ is the probability of outcome $+1$ when measuring ρ in the kth PO. Given these considerations, the cost function used is

$$\mathscr{L}(\mathbf{v}) = N_s(\mathbf{v}) \langle S'(p_k(\mathbf{v})) \rangle_{k \in K}, \quad (19)$$

where K is the subset of $\lceil 3^n f^e \rceil$ POs (index k) with smallest $S'(p_k)$. This way, the cost function is directly proportional to the size N_s of the set of states, and is only affected by the $\lceil 3^n f^e \rceil$ best-matching POs, as they allow the encoding to reach fidelity f^e. To minimize the cost function, we flip individual bits of $\mathbf{v}$ and choose the change that decreases the cost function the most, introducing occasional random bit flips to avoid getting stuck in local minima. The exact procedure is detailed in Algorithm 1.

Applying the eigenstate selection to our example, we obtain perfect encoding of our logical bitstring with as few as $N_s = 14$ states, and a SR for parity bit retrieval as low as 68 for lower fidelity values, as can be seen in Table II.

Set size (N_s)	PPC fidelity (f^e)	Chosen ν	SR (S)
9	75/81	0.007	68
11	76/81	0.006	88
12	78/81	0.004	103
13	80/81	0.001	135
14	81/81	0.001	147

TABLE II. Fidelity and SR to encode and retrieve a high-RC PPC of an arbitrary logical bitstring, using N_s context eigenstates of 4 qubits. Complete fidelity is obtained for $N_s = 14$, and $N_s = 9$ is a lower bound to set sizes that can span all POs. The SR is calculated with $\nu \ll (1 - f^e)$ such that $f \approx f^e$, excepting the $N_s = 14$ case where necessarily $\nu > (1 - f^e)$ (but $f \approx f^e$ still holds).

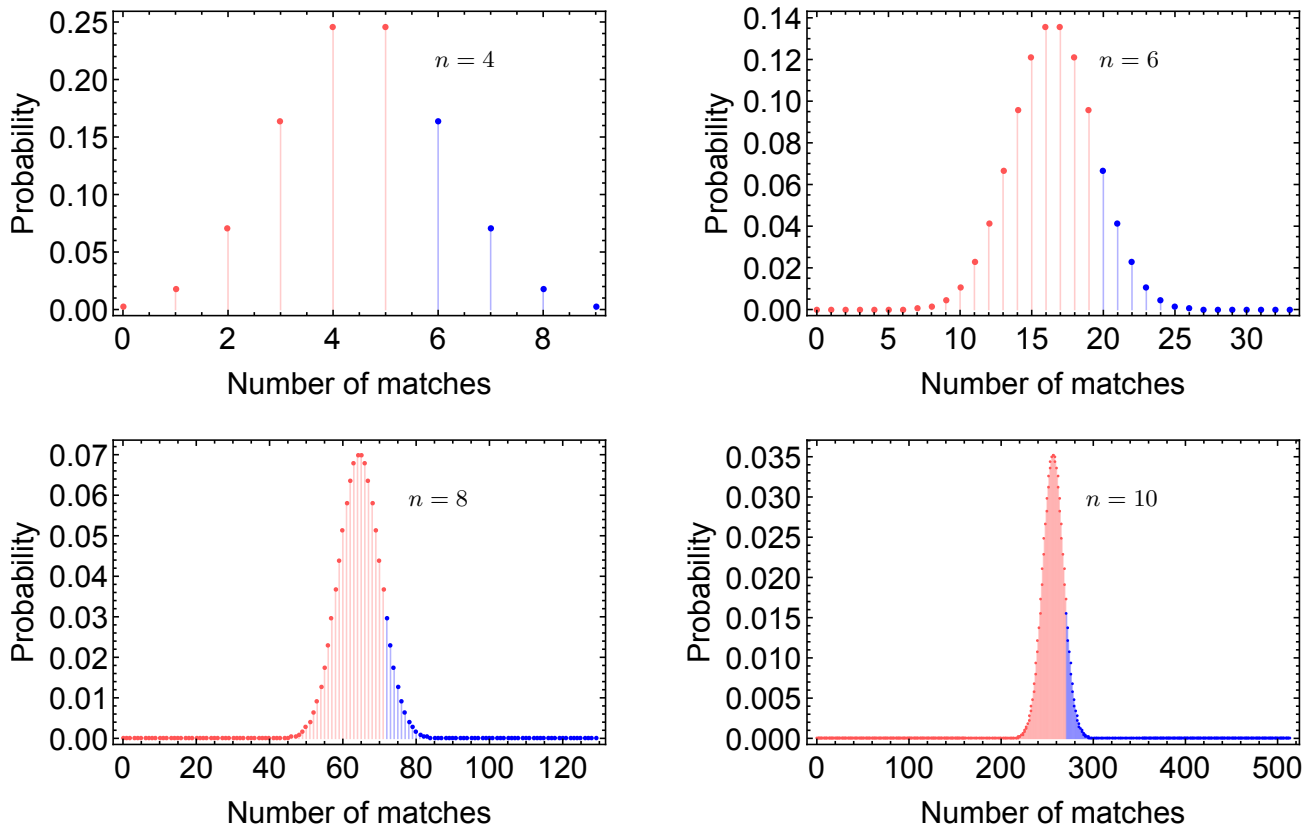

FIG. 6. Probability distribution for the number of matches between a context eigenstate of n qubits and an arbitrary PPC for the relevant POs, for contexts of maximum size $d_{\mathcal{C}} = 2^{n-1} + 1$, and $n = 4, 6, 8, 10$. The number of matches have been modeled with binomial distributions $B(d_{\mathcal{C}}, 1/2; \cdot)$, and the upper 10th percentile of matches $n_m | \sum_{i=n_m}^{d_{\mathcal{C}}} B(d_{\mathcal{C}}, 1/2; i) \leq 1/10$ has been colored differently.

Binomial modelling the matches

The number of matches between a context eigenstate (for contexts of maximum size $d_{\mathcal{C}} = 2^{n-1} + 1$) and an arbitrary PPC bitstring (of size $d_{\mathcal{C}}$) for that context can be modelled by a binomial distribution $B(d_{\mathcal{C}}, 1/2; \cdot)$ (Fig. 6), i.e. each parity has probability 1/2 of matching the eigenstate and we can expect half the parities to match. As n increases, the distribution for the number of matches becomes sharper, which increases the precision of statistical approaches making use of expectation values.

To encode an arbitrary PPC, we search for context eigenstates from among all contexts and select those with the most matches. This selection must be large enough to guarantee that all POs appear a reasonable number of times, and also small enough for them to be a selected group with high number of matches. Hence we define an encoding error tolerance ε which serves as a cutoff to select states on the higher percentile of matches.

Protocol repetitions

To improve the fidelity and optimize the error tolerance, we perform an odd number r of repetitions of the full protocol (encoding and decoding) fixing each time a different PO ordering $\{\mathcal{G}\}$, and thus allowing for independent statistics. Although this procedure increases the total SR by a factor of r, $S_r = rS$, it increases considerably the final logical bit fidelity, defined as the probability that the majority of repetitions yield the correct logical bit,

$$\bar{f}_r = \sum_{j=(r+1)/2}^{r} B(r, \bar{f}; j). \tag{20}$$

The optimal encoding error $\varepsilon^\star$ is obtained by choosing a number r so that $\bar{f}_r$ is approximately constant for various values of ε, and then minimize $S_r(n, \varepsilon)$. For large n and $0.95 \leq \bar{f}_r \leq 0.999$ taking steps of 0.001, this yields $\varepsilon^\star = 0.0480$ (± 0.0034), with a low dependence on $\bar{f}_r$, as can be seen from the small standard deviation.

We then compute $N_{\mathrm{s}}(n, \varepsilon^\star) = (3/2)^n / g(\varepsilon^\star)$ with $g(\varepsilon^\star)^{-1} = 1.385$ from Eq. (3), and determine $f^{\mathrm{e}}(n, \varepsilon^\star) =$

0.722 and $\bar{f}^{\mathrm{e}}(n, \varepsilon^\star) = 0.598$ from Eq. (4) (for large n). This implies that for low retrieval noise (such that $\bar{f} \approx \bar{f}^{\mathrm{e}}$), $r = \{69, 137, 239\}$ repetitions are sufficient to achieve $\bar{f}_r = \{0.95, 0.99, 0.999\}$ fidelity, respectively. All this considered, we also compute r as a function of $\bar{f}_r$, to properly express the SR as a function of the fidelity and the quantum system size. This is done isolating r in Eq. (20) by taking the Normal approximation, $B(r, \bar{f}; \cdot) \to N(\mu = r\bar{f}, \sigma^2 = r\bar{f}(1-\bar{f}))$, and considering that $\mathrm{Erfc}^{-1}(y) \approx \sqrt{\ln(1/y)}$ for $y \ll 1$. This way, $r \approx 50 \ln(1/(2 - 2\bar{f}_r))$.