High Fidelity Rydberg Gates: A Review of Recent Techniques

Will Staples and Siddhant Midha
Princeton Quantum Initiative, Princeton University, NJ USA 08544
(Dated: December 13, 2024)

We review the recent developments in high-fidelity entangling gates using Rydberg interactions of neutral atom qubits. We focus on the work by Evered et al. [1] realizing two-qubit phase gates with 99.5% fidelity on up to 60 atoms in parallel. We provide an overview of key techniques, including optimal control methods, atomic dark states, and various benchmarking strategies. We report the key results and provide an outlook for future progress and applications.

I. INTRODUCTION

Neutral atom quantum processors have emerged as one of the most promising platforms in quantum technology, gaining traction in quantum information processing [2] as well as quantum simulation of complex many-body dynamics [3]. Hyperfine and metastable states within trapped ions are a natural choice for the storage of quantum information because of their exceptionally long coherence times [4]. Single atom initialization, local addressing and high-fidelity readout have been demonstrated with multiple atomic species [5–7]. The strong, long-range interactions via by the Rydberg blockade mechanism [3] enables the implementation of multi-qubit gates. These interactions can be turned on and off with a contrast of 12 orders of magnitude, enabling high fidelity quantum logic [8, 9]. Additionally, atomic qubits offer spatial mobility, which enables high system connectivity. Freedom from locality constraints allows for enhanced flexibility, particularly when considering encoding information within an error-correcting code.

High-fidelity entangling gates is a major thrust within the field of quantum computing. Specifically, it is essential to get error rates *below threshold* [10] to enable large-scale fault-tolerant quantum computation [11]. Herein, the outstanding challenge within neutral atom computing has been improving the two-qubit fidelity above and beyond the previous demonstrations of 97.5%, and bridging the gap with the state-of-the-art entangling operations in other platforms such as superconducting qubits (99.4%, 74 qubits) [12] and ion-trap computers (99.4 – 99.6%, 30 qubits) [13]. This gap been closed in recent years with 99.4 – 99.7% fidelity achieved [1, 14, 15] in multiple experiments.

In this paper, we examine the work of Evered et al. [1], starting with a detailed discussion of their approach to qubit manipulation. We then review prior techniques for Rydberg entangling gates, followed by an in-depth exploration of the optimal control techniques employed to achieve high-fidelity gates using the Rydberg blockade mechanism. We also discuss the benchmarking methods used to evaluate these gates. After presenting the results from [1], we conclude with a segue to the concept of the elusive *path of three nines*.

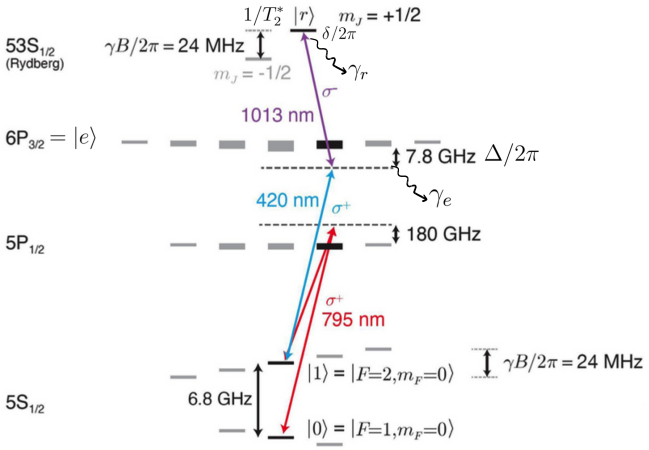

FIG. 1. Level Diagram of ^{87}Rb used in [1]. Shown are the two computational states $\{|0\rangle, |1\rangle\}$, which couple to the $5P_{1/2}$ state for single-qubit gates via Raman transitions at 795nm. The two-photon Rydberg transition between $|1\rangle$ and $|e\rangle = 6P_{3/2}$ is driven via 420nm and 1013nm lasers. The level-splitting induced by a constant DC magnetic field of 8.5G is also depicted. The main error sources shown: Rydberg-state decay γ_r, intermediate-state scattering γ_e, and Rydberg dephasing T_2^*. Two qubit gates involve modulating the Rabi frequency $\Omega(t)$ and phase $\phi(t)$ profiles of the 420nm leg of the two-photon transition.

A. The Qubit

Qubits in this work are stored in the hyperfine structure of trapped ^{87}Rb atoms. These atoms are individually trapped in optical tweezers and cooled to near the motional ground state (10μK, with estimated radial phonon number $\bar{n} \approx 1 - 2$). This set-up allows them to be moved precisely, which will be a critical component of entangling gates.

The computational basis states are given by the clock states in the $m_F = 0$ hyperfine manifold, viz., $|0\rangle := |F = 1,\ m_F = 0\rangle$ and $|1\rangle := |F = 2,\ m_F = 0\rangle$, split by a spatially uniform magnetic field. Outside of the computational subspace, the highly energetic $53S_{1/2}$ Rydberg state $|r\rangle$ is used for entangling operations (the details of which are the concern of the bulk of this work). The $|r\rangle$ state

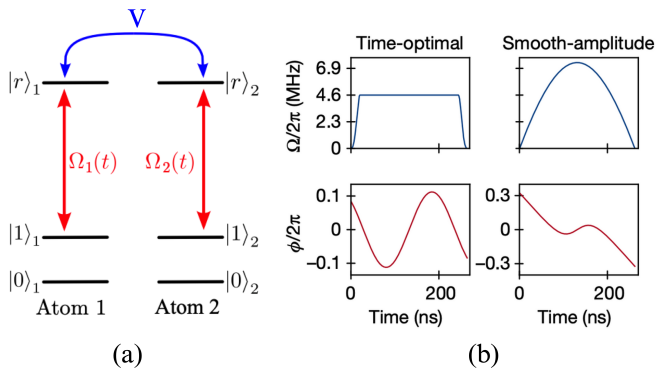

FIG. 2. (a) Simplified level diagram for two-atom interaction via the Rydberg blockade mechanism [16]. (b) Pulse profiles of the time-optimal and smooth-amplitude gates demonstrated in [1].

is accessed via a two-photon Raman transition with the $6P_{3/2}$ as an intermediate excited state, denoted $|e\rangle$ (see Fig. 1).

Initialization proceeds via a pumping scheme to the $|0\rangle$ state. The qubit is first pumped into the $|F=2, m_F=2\rangle$ state with σ^+ polarized 780nm light. Then, the magnetic field is rotated by 90^o such that the Raman laser propagation axis has an orthogonal component and can drive $\sigma^\pm$ transitions in the hyperfine manifold. Further, two separate Raman pulses are applied which transfer population from $|F=2, m_F=2\rangle$ to $|F=1, m_F=1\rangle$ and then to $|F=2, m_F=0\rangle$. The magnetic field is then rotated back. The estimated pumping fidelity is 99.7 – 99.8%. Measurements are performed by shining light to strongly couple $F=2$ to $F'=3$ on the D2 transition, which heats up atoms in $|1\rangle$ and subsequently expels all atoms in the $|1\rangle$ state from the traps. This, in-turn is a fluorescence based readout of the atom array with combined pushout and imaging fidelity of 99.83%.

Single-qubit gates are executed using Raman transitions blue-detuned from the $5P_{1/2}$ intermediate state. These have high fidelity (99.97% as measured by single-qubit randomized benchmarking). The fidelity of these gates is primarily limited by the population of the $5P_{1/2}$ state. Since this scales with Ω/Δ, error could be further suppressed by increasing the detuning Δ. However, because this is still more than an order of magnitude better than the improved two-qubit gate performance, single-qubit errors aren't a significant concern.

B. Previous Entangling Gates

Prior to the work presented here, state of the art neutral atom processors used entangling gates mediated by a global laser (or pair of lasers) with a discrete phase jump [7]. This laser is resonant with the transition between the $|1\rangle$ state and the excited $|r\rangle$ state. This gate should be understood by examining the dynamics of the three states of interest: $|00\rangle$, $|01\rangle$, and $|11\rangle$.

Firstly, the laser is not resonant with $|0\rangle$, so the gate acts as the identity on $|00\rangle$. Secondly, $|01\rangle$ is driven in coherent oscillations with the state $|0r\rangle$. As is the familiar for two-level systems, the frequency of this oscillation is $|\Omega|$ where Ω is the coupling strength. Finally, the most complicated case is the dynamics of $|11\rangle$. To understand this we examine the effective Hamiltonian of the 4-level subspace $\{|11\rangle, |1r\rangle, |r1\rangle, |rr\rangle\}$. The Hamiltonian for a two-atom interaction in the rotating frame reads (see Fig. 2(a)),

$$H = \frac{\Omega_1}{2}|1\rangle_1\langle r|_1 \otimes \mathbb{1} + \mathbb{1} \otimes \frac{\Omega_2}{2}|1\rangle_2\langle r|_2 + V|rr\rangle\langle rr| + h.c. \quad (1)$$

Since we are concerned with the case of *global* control, with $\Omega_1 = \Omega_2 \equiv \Omega$, and now the Hamiltonian matrix is,

$$H_{11} = \frac{1}{2}\begin{pmatrix} 0 & \Omega & \Omega & 0 \\ \Omega^* & 0 & 0 & \Omega \\ \Omega^* & 0 & 0 & \Omega \\ 0 & \Omega^* & \Omega^* & V \end{pmatrix} \quad (2)$$

Where V is the so-called "Rydberg Blockade" [3] which results from the Van-der-Waals interaction of nearby atoms in the $|r\rangle$ states. For the purposes of understanding the gate, the fact that $V \gg |\Omega|$ means that Hamiltonian is approximately block diagonal, and the $|rr\rangle$ state is not populated. Note that this in-turn implies a higher $|\Omega|$ can potentially induced finite-blockade effects. We may further block-diagonalize the remaining 3-state Hamiltonian by changing our basis to $|11\rangle$, $(|1r\rangle \pm |r1\rangle)/\sqrt{2}$. This change of basis gives us:

$$\begin{aligned} H_{11} &= \frac{1}{4}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 0 & \Omega & \Omega \\ \Omega^* & 0 & 0 \\ \Omega^* & 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \frac{\sqrt{2}}{2}\begin{pmatrix} 0 & \Omega & 0 \\ \Omega^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned} \quad (3)$$

This makes an effective two-level system which oscillates between $|11\rangle$ and a state $|W\rangle := (|1r\rangle + |r1\rangle)/\sqrt{2}$ with frequency $\sqrt{2}|\Omega|$. Since $|W\rangle$ is outside the computational subspace, our gate should return $|11\rangle$ to itself with some phase. Since we also need to return $|01\rangle$ to itself up to a phase, this cannot be accomplished by any single rotation.

A simple solution to this problem is to rotate $|11\rangle$ for a full rotation while partially rotating $|01\rangle$. We can then switch the axis of rotation and do another full period for the $|11\rangle$ state. If the axes are chosen well, this can also return $|01\rangle$ to itself (Fig. 3). The states $|01\rangle$ and $|11\rangle$ return with phases ϕ_{01} and ϕ_{11} respectively.

Changing the axis of rotation is accomplished by changing the phase $\phi = \arg(\Omega)$ of the laser driving the transition. An equivalent picture of this task is changing the detuning Δ of the laser, which is related to the phase by $\Delta = -d\phi/dt$ [1, 16].

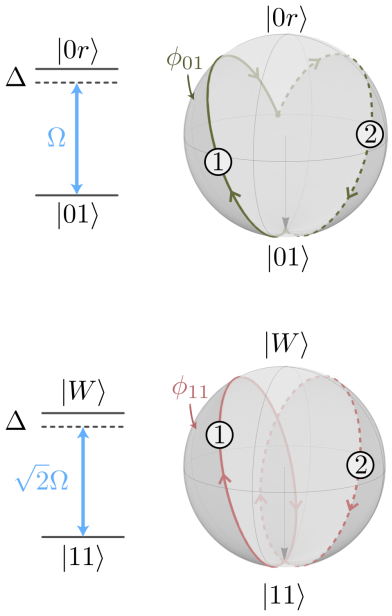

FIG. 3. The dynamics of the $|01\rangle$ and $|11\rangle$ states during the global gate pulse can be modeled as two level systems with excited states $|0r\rangle$ and $|W\rangle$ respectively. The Rabi frequencies of these two systems differ by a factor $\sqrt{2}$, so the computational subspace cannot be recovered while rotating about a time-independent axis. Switching between two axes allows recovery while applying phases ϕ_{01} and ϕ_{11} on each state. [7]

There are many possible sequences of two pulses separated by a phase jump which satisfy the condition of mapping the computational subspace to itself. Each results in a different relative phase $\phi_{11} - 2\phi_{01}$. There is however, a unique solution which gives a relative phase π, which corresponds to controlled Z gate (up to local unitaries).

II. METHODS

A. Optimized Gate Implementations

The control parameter in designing the gates for this system is effectively $\Omega(t)e^{\iota\phi(t)}$, giving us real functions of time— Ω and ϕ— to be optimized for achieving fast, high-fidelity entangling gates. More precisely, if $U(T) := \mathcal{T}\exp\left\{-\iota\int_0^t H(\tau)d\tau\right\}$ is the unitary implemented by the Hamiltonian evolution and U_0 is the desired unitary (for instance, the CZ or CCZ gates) then we wish to maximize the average gate fidelity, viz.,

$$F = \int_{\psi\in\mathcal{H}^{\otimes 2}} d\psi |\langle\psi|U_0^\dagger U(T)|\psi\rangle|^2 \tag{4}$$

where the integral is over the two-qubit Hilbert space $\mathcal{H}^{\otimes 2}$. Optimal control methods are used to achieve this task, leading to time-optimal [1, 16] and smooth-amplitude pulse sequences [1], as we discuss now.

1. Time-Optimal Gates

Entangling gates on neutral atoms have the following general sources of error [16]:

1. Scattering from intermediate states of two-photon transitions
2. Finite lifetime of Rydberg state
3. Variation in Rydberg blockade strength (coupled through position)
4. Laser phase and intensity noise
5. Doppler shifts (thermal motion)

Many of these errors, particularly the first two, scale with the duration of the gate. Because of this, optimizing for minimal gate duration is a reasonable proxy for fidelity maximization.

This idea can be used in conjunction with Gradient Ascent Pulse Engineering (GRAPE), which is a computational technique used for optimal quantum control [17]. Using GRAPE requires specifying an initial state ψ_0, a gate duration T, a parameterized unitary $U_\alpha(t)$, and a cost function $\mathcal{J}(\psi)$. Then it finds the parameter vector α that minimizes $\mathcal{J}(U_\alpha(T)\psi_0)$.

In the context of implementing the CZ gate with only global control (which is desirable for experimental feasibility), the unitary must act symmetrically on $|01\rangle$ and $|10\rangle$. Because of this, the unitary is completely defined by its action on the initial state $\psi_0 = |01\rangle + |11\rangle$. This feature makes analysis with GRAPE relatively straightforward [16]. We can define a cost-function $\mathcal{J} = 1 - F$ with the fidelity F given by:

$$\begin{aligned} F &= \frac{1}{20}\left(|1 + 2a_{01} + a_{11}|^2 + 1 + 2|a_{01}|^2 + |a_{11}|^2\right) \\ a_q &= e^{\iota\xi_q}\langle q|U_\alpha(T)|\psi_0\rangle \end{aligned} \tag{5}$$

All that is left is to optimize over $\Omega(t)$. If make the physically reasonable assumption that we have access to only values of Ω with $|\Omega| \le \Omega_{\max}$, then the time optimal gate must have $|\Omega(t)| = \Omega_{\max}$ for all t. If this were not the case, the gate could be accelerated by simply increasing $|\Omega(t)|$ at sub-maximal values of t and compressing $\phi(t) = \arg(\Omega(t))$ appropriately. We then optimize over $\phi(t)$. Since the time derivative of phase is detuning, this function must be differentiable. With this condition, we can discretize ϕ in a physically sensible way.

Applying GRAPE for many values of the unitless parameter $\Omega_{\max}T$ shows that there is a threshold value $\Omega_{\max}T_0$ which is the minimum possible value for a high fidelity gate (Fig. 4). The pulse shown assumes that the Rydberg block V is arbitrarily large and thus $|rr\rangle$ is completely unpopulated throughout the duration of the gate. However, this assumption is not necessary for this numerical method to work; qualitatively similar pulses can be found for any finite value of V/Ω. In physical systems, V

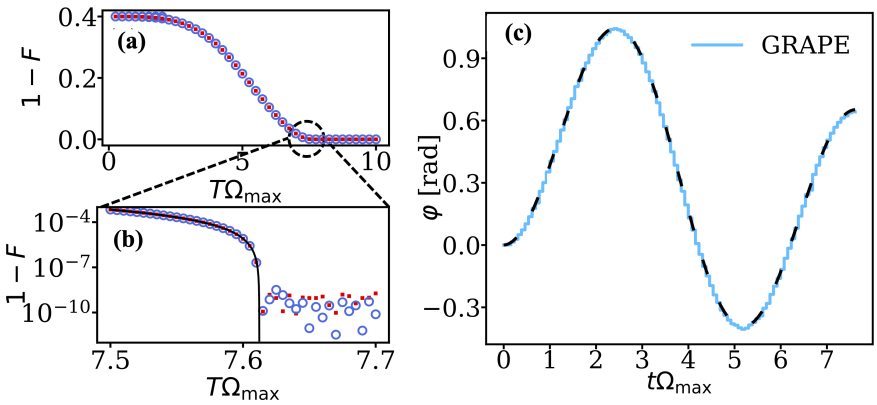

FIG. 4. (a) Maximum possible fidelity found by GRAPE for different values of gate duration T. This is normalized by maximum available Rabi frequency $\Omega_{\max}$. (b) Higher resolution plot on a log scale of the same as (a). We see that GRAPE fails to increase fidelity when loosening duration constraints past some critical duration $T_*\Omega_{\max} = 7.612$. Increasing the parameter dimension available to GRAPE does not significantly alter this behavior. (c) Phase pulse proposed by GRAPE for gate duration $T = T_*$. [16]

is a function of interatomic distance. We can then make an ideal gate for a fixed distance, but variation from this distance is a source of infidelity.

The fixed amplitude, or time optimal gate employed in [1] is thus of the form (see Fig. 2(b)),

$$\phi(t) = A\cos(\omega t - \varphi_0) \tag{6}$$

constituting an exact gate with $\Omega T/2\pi = 1.215$, thus being 0.2% slower than the gate in [16].

2. *Smooth Amplitude Gates*

In an effort to further generalize the idea of a time-optimal gate proposed by Jandura and Pupillo [16], Evered et al. [1] also propose a pulse which has time dependent $|\Omega(t)|$. The purpose of this gate design was to further suppress scattering from the intermediate state $|e\rangle$. Based on the non-Hermitian Hamiltonian which includes the term $\iota\gamma_e |e\rangle\langle e|$ (see Appendix A) to capture intermediate scattering, they were able to infer the functional form (see Fig. 2(b)),

$$\begin{aligned} \frac{\Omega_{420}(t)}{\Omega_{1013}} &= \Omega_0 + \Omega_1 \mathrm{sech}(\omega\tau)^\kappa \\ \phi(t) &= \delta\tau + B\tanh(\lambda\tau) \end{aligned} \tag{7}$$

They provide no analytical justification for this form, but rather seem to use qualitative analysis [1]. The pulse can then be optimized over the parameters in the above form using GRAPE. While this gate suppresses intermediate-state scattering and reduces off-resonant coupling to other states it is more susceptible to finite-blockade effects owing to larger maximal Rabi frequencies during the gate profile. [2] Moreover, the time-optimal gate is actually slower ($\Omega T/2\pi = 1.215$) than the smooth amplitude gate ($\Omega T/2\pi = 1.207$).

We remind the reader[3] that the amplitude and phase of the 1013nm laser is kept constant at all times, and the two photon detuning δ is effectively time dependent as $\delta := \delta(t) \propto -\phi'(t)$. Moreover, $\Omega \equiv \Omega_{420}\Omega_{1013}/2\Delta$.

We can develop a qualitative understanding of how this gate suppresses intermediate state scattering by considering the dark-state basis for the two-photon transition $|1\rangle \to |e\rangle \to |r\rangle$. If we consider the intermediate detuning Δ as the dominant energy scale (as compared with the Rabi frequencies and the two-photon detuning δ), we can write this basis to first order in Ω_{1013}/Δ as:

$$\begin{aligned} |D\rangle &:= \frac{-1}{\sqrt{1+\alpha^2}}|1\rangle + \frac{\alpha}{\sqrt{1+\alpha^2}}|r\rangle \\ |B\rangle &:= \frac{\alpha}{\sqrt{1+\alpha^2}}|1\rangle + \frac{\Omega_{1013}\sqrt{1+\alpha^2}}{2\Delta}|e\rangle + \frac{\alpha}{\sqrt{1+\alpha^2}}|r\rangle \\ |E\rangle &:= \frac{-\alpha\Omega_{1013}}{2\Delta}|1\rangle + |e\rangle - \frac{\Omega_{1013}}{2\Delta}|r\rangle \end{aligned} \tag{8}$$

with $\alpha \equiv \alpha(t) := \Omega_{420}(t)/\Omega_{1013}$. This basis allows us to consider the gate as a nearly perfect two-level system, since population of $|E\rangle$ is suppressed by $(\Omega_{1013}^2\delta^2/\Delta^4)$. The task of minimizing intermediate state scattering can then be reduced to minimizing the population of the bright state $|B\rangle$ during the execution of the gate, since this corresponds directly with the population of $|e\rangle$.

[1] Analytical justification for this form eludes us. It seems this was pulled out from their collective caboose.

[2] We are unsure why they don't use that maximal Rabi frequency for the time-optimal gate.

[3] Genyue and/or Jeff

All time-scales in the gate, including the rise time ($\approx 10ns$) of the blue laser in the time-optimal gate, are adiabatic with respect to the splitting Δ. Thus, upon turning on the blue laser, the dressed initial state is

$$|1\rangle = |1\rangle + \frac{\alpha\Omega}{2\Delta}|e\rangle \approx \frac{\alpha}{\sqrt{1+\alpha^2}}|B\rangle - \frac{1}{\sqrt{1+\alpha^2}}|D\rangle \quad (9)$$

Note that the $|E\rangle$ is the highest energy state, separated from the $\{|B\rangle, |D\rangle\}$ manifold by Δ. We can thus describe the evolution of $|1\rangle$ upon turning on the red laser in the $\{|B\rangle, |D\rangle\}$ subspace with the following effective Hamiltonian

$$-\delta/2(|B\rangle\langle D| + |D\rangle\langle B|) + \frac{\Omega_{1013}^2}{2\Delta}|D\rangle\langle D| \quad (10)$$

Now, upon choosing $\Delta<0$ and $\delta>0$ we increase population in the dark state, viz., $P_D - P_B \sim |\delta\Delta|\Omega_{1013}^2/(2\delta^2\Delta^2 + \Omega_{1013}^4)$. Moreover, the time-optimal gate is operated at $\alpha(t) = 1$, whereby the smooth-amplitude gate starts from $\alpha \ll 1$.

However, δ is of the same order of magnitude as Ω, which is roughly 3 orders of magnitude smaller than Δ. Because of this, true adiabatic transfer from $|1\rangle$ to $|D\rangle$ is not possible without drastically increasing gate-time. Phrased another way, one simply cannot operate at $\alpha \ll 1$ for the entire gate profile as a larger admixture of the bright state is needed to realize the Rydberg population. Nonetheless, we can intuitively understand that gradually turning the gate on and off will suppress scattering in comparison to the time-optimal approach.

The true picture is more complicated than the intuitive picture we have described above. Nonetheless, Evered et al. [1] claim that their numerics support the basic conclusion that the degree of possible scattering suppression increases as they loosen the restrictions on gate-time. They note that a gate the same duration as the time-optimal [4] gate can suppress scattering by a factor of 1.2 and that a gate twice as long as time-optimal has a suppression factor of 2.5.

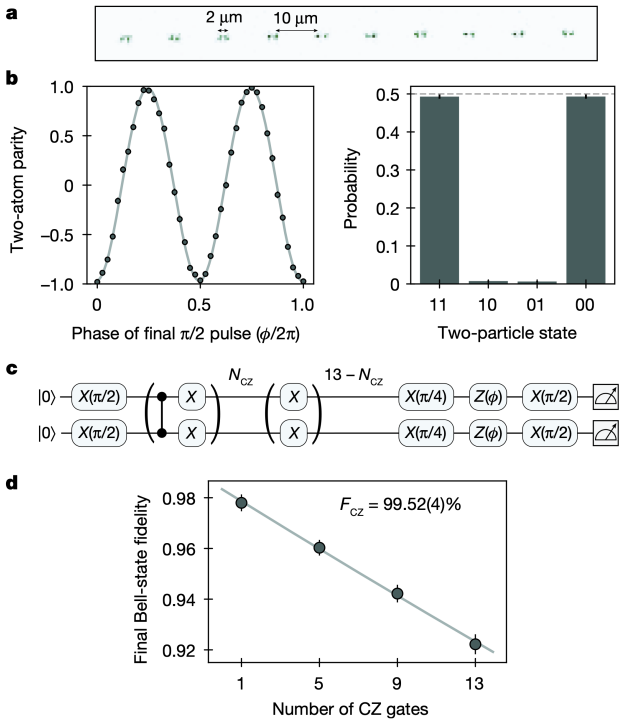

FIG. 5. (a) Spacing for parallel implementation of 10 CZ. (b) Example plot for the extraction of Bell-state fidelity. This particular plot corresponds to raw Bell-state fidelity after a single CZ gate. (c) Circuit for measuring train of CZ gates. (d) Plot illustrating fidelity with variable number of CZ gates along with an exponential fit. This shows the time-optimal protocol[1]

B. Benchmarking

The term *randomized benchmarking* (RB) [18] describes a set of protocols and tools widely used within the community to characterize noisy quantum devices. The goal is to estimate the errors of noisy devices in a fashion that is immune to state preparation and measurement (SPAM) faults. The general paradigm of RB falls into the following two parts [19]:

1. *data-collection.* (1) prepare a (ensemble of) states of the device (2) apply a sequence of random quantum operators (sequence length T) (3) apply a correction unitary, and (4) measure

2. *postprocessing.* (1) fit the measurement histogram to a known functional form and (2) extract relevant fidelities

More concretely, if the sequence of random unitaries is $g_1, g_2 \ldots, g_T$, the the correction unitary is $g_c := g_1^{-1} g_2^{-1} \ldots g_T^{-1}$ and brings back the qubit state into some fiducial state in the noise-free case. With noise, one can generally (see [19] for rigorous guarantees) fit a form of constant $\times F^t$ to the return probability. Thus, experimentally measuring the return probability as a function of the sequence length t provides a good estimate of the fidelity F. Full RB typically requires extensive experimental control as the set of gates applied has to be random enough. Specifically for estimating 2-qubit Rydberg CZ gate fidelities, one can consider the following two situations [15],

- (True fidelity)
  $F_{\text{Haar}} = \mathbb{E}_{\psi\sim\text{Haar}}\text{Tr}[CZ|\psi\rangle\langle\psi|CZ\,\mathcal{U}_{CZ}(|\psi\rangle\langle\psi|)]$
- (Symmetry-restricted fidelity)
  $F_{\text{sym}} = \mathbb{E}_{\psi\sim\text{Sym-Haar}(2)}\text{Tr}[CZ|\psi\rangle\langle\psi|CZ\,\mathcal{U}_{CZ}(|\psi\rangle\langle\psi|)]$

where $\mathcal{U}_{CZ}$ is the actual (possibly non-unitary) operation implemented and the latter includes an average over all the states which are symmetric under a permutation of

[4] Again, the name "time-optimal" appears to be a smidgen of a misnomer.

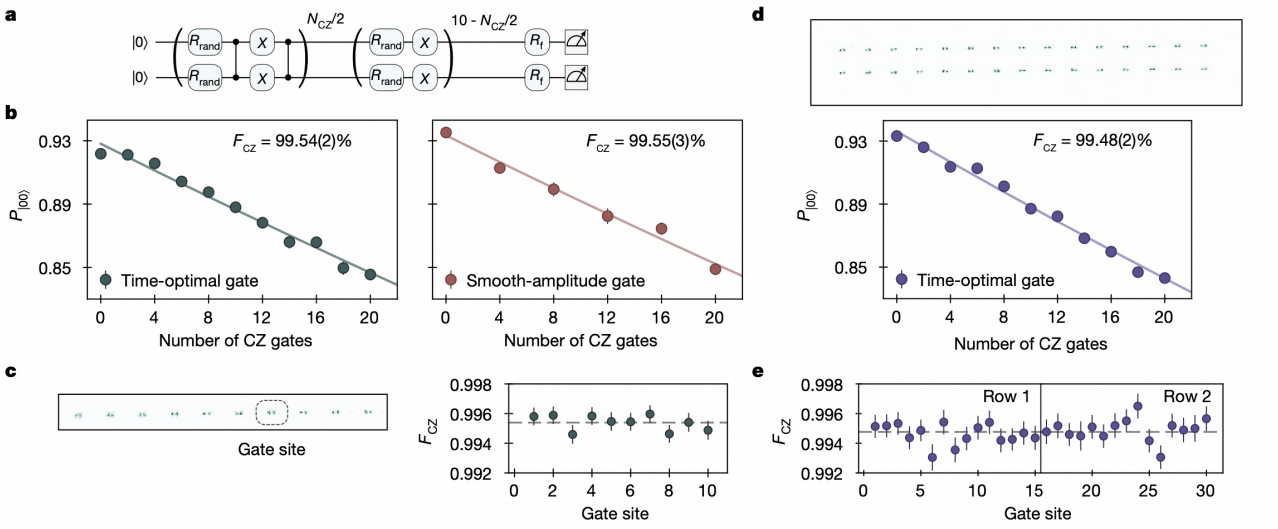

FIG. 6. (a) Circuit for global RB, $R_{rand} \sim$ Haar(1). (b) RB results on 20 atoms in parallel as a function of sequence length. (c) Statistical deviations across the atom array. (d-e) RB on 60 atoms in parallel with statistical deviations across the array. [1]

the two qubits. The need for the latter naturally arises in the Rydberg scenario due to the global controls.

III. RESULTS

A. Characterization of CZ Gates

Evered et al. [1] characterize their entangling gates by measuring the quality of the Bell states they produce. More explicitly, they create the Bell states $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and apply a global $\pi/2$-pulse with a Z variable phase. The authors then construct the correlator $\langle Z_1 Z_2 \rangle$ from transversal Z basis measurement. For an ideal state $|\Phi^+\rangle$, this correlator will oscillate between ± 1. An imperfect Bell state will result in oscillation with amplitude less than 1. This amplitude is a good approximation of the fidelity of your procedure [20]. We study the Bell fidelity a bit more in depth in Appendix B[5].

Using this technique, they measure a raw Bell-state fidelity of 98.0(2)%. However, the infidelity of these states is dominated by SPAM errors. They account for these by preparing and subsequently measuring computational basis states and estimating the probability of atom loss during the gate. The result is a corrected Bell state fidelity of 99.4(4)%.

Additionally, they more systematically extract the gate infidelity by applying odd-length trains of CZ gates and measuring the fidelity of the resultant Bell pair. By fitting an exponential curve to the fidelity with respect to the number of CZ gates used, the marginal contribution of each gate can be inferred quantitatively Fig. 5. This finds a two-qubit entangling gate fidelity of $F_{\mathrm{CZ}} = 99.54(2)\%$.

[5] we might have gotten a little obsessed

B. Global Randomized Benchmarking

In yet another fashion of interleaved RBs, Evered et al. [1] demonstrates the high-fidelity operation of the two-qubit CZ gate via interleaving CZ gates with global single qubit Haar random gates. This is shown in Fig. 6(a). We outline subtleties with this 'global' part of this method in Sec. III D. Similar to the general RB protocol discussed before, a precomputed rotation at the end of the sequence R_f returns the state to the determined $|00\rangle$ state.

The return probabilities for a parallel operations on 20 atoms in a linear array (see Fig. 6(c)) is shown for the time-optimal and smooth-amplitude gates in Fig. 6(b). The authors extract a fidelity of $F_{CZ} = 99.54(2)\%$ and $99.55(3)\%$ respectively. This shows good agreement with the two methods of optimized Rydberg gates discussed. Shown further in Fig. 6(c) is the deviations in fidelity from this mean value throughout the 1-dimensional chain, showing good homogeneity and suggesting scalability of such global control methods.

To further exemplify scalability, the authors implement the same procedure in parallel on 60 atoms in a chain, which are shown in Fig. 6(d-e). The fidelity extracted from an exponential fit is 99.48(2)%, in good agreement with the previous numbers. Furthermore, the spatial deviation profile of Fig. 6(e) shows a homogenous implementation. This enables the authors to make a statement about the methods presented already being applicable to hundreds of qubits via shuttling atoms in and out of the entangling zone during a circuit. Moreover, further improvements demand continually scaling laser

power and maintaining beam homogeneity, as other microscopic decoherence mechanisms seem to indicate an $O(1)$ contribution to the net infidelity as the array size increases.

C. Native Multi-Qubit Gates

By design, we have focused the bulk of this review on two-qubit entangling gates, giving minimal attention to entangling gates involving 3 or more qubits. This is because the principles and methods we've discussed in the context of two-qubit gates can be extended to multi-qubit gates. Evered et al. use the optimization techniques described in our Methods section to design a global pulse for implementing CCZ gate. The pulse is time-optimal (Ω has a "top-hat" amplitude profile with oscillating phase) and only 44% longer than their time-optimal implementation of CZ. It's worth noting that a similar pulse was proposed by Jandura and Pupillo [16].

In analogy with the measurement of Bell-state fidelity, they use these CCZ gates to prepare three-qubit Greenberger-Horne-Zelinger (GHZ) state and subsequently measuring fidelity. Similarly to their characterization of CZ, they prepare a train of these CCZ gates and fit an exponential to the relation between fidelity and number of gates. This procedure yields a three-qubit entangling gate fidelity of $F_{\mathrm{CCZ}} = 97.9(2)\%$.

Implementing a CCZ gate requires a minimum of 6 two-qubit entangling gates. As a consequence, the gate proposed by Evered et al. is not only far faster than its equivalent construction from one and two-qubit unitaries, it is also higher-fidelity. This suggests that native multi-qubit gates as a promising avenue.

D. Digging into the Randomized Benchmarking

Ref. [15] points out a protocol for RB which (1) takes into account the global control of the RB setup and (2) reduces sensitivity to single-qubit errors. Thus, the goal of such a protocol is to faithfully extract the *Rydberg gate* error, and isolate fully the single qubit error. This is important, as while the 'global RB' protocol of [1] that we outlined approximates F_{sym}, it is sensitive to single qubit errors as the distribution over two qubit states varies as a function of sequence length. Whereas, the protocol should be averaged over all two qubit (Clifford) states. Thus, if the time spent in an entangled state increases with the CZ length sequence, errors due to single qubit imperfections accumulate as the sequence increases.

Thus, the symmetric stabilizer benchmarking (SSB) method of [15] accounts for this via noting that the symmetric subspace of two-qubit Clifford states has twelve elements, and then starting with a initialization unitary U_{init} which prepares these 12 states uniformly randomly. Then, CZ gates are interleaved with single-qubit *global* random gates from the ensemble $\{R_{X/Y(\pm\pi/2)}\}$ (thus, similar to [1], but Clifford-random single-qubit instead of Haar-random). Finally, a recovery unitary is applied to return to a predetermined state. With this, [15] proposes modifications to the circuit of Evered et al. to successfully reduce single qubit sensitivity. Detailed quantitative discussions can be found in ([15] Appendix. E).

IV. CONCLUSIONS AND OUTLOOK

In this paper, we studied the problem of two-qubit gates using the Rydberg blockade. Focusing on the recent work by Evered et al., we discussed the hyperfine ^{87}Rb qubit, and studied the mechanism behind the two-qubit CZ gate via Rydberg blockade. Further, we surveyed optimal control techniques used to achieve high fidelity CZ gates, and discussed the benchmarking protocols. Finally, we detailed in depth the Bell-state benchmarking and global randomized benchmarking presented in [1]. Post a brief review of the native Rydberg multi qubit gates, we contrasted the subtleties in the RB protocol of [1] with the recent points outlined in [15]. Now, we end this work with a brief discussion of the path to 0.999, and relevance to the crucial goal of below-threshold error correction.

A. Path to Three Nines

Ref. [15] points out that since infidelity due to Rydberg decay scales as Ω^{-1} and finite atomic temperature as Ω^{-2}, higher Rabi frequencies are key in the step towards three nines. Quantitative calculations using the Fidelity Response Theory (FRT) developed in [15] predicts that cavity filtering is likely to be very beneficial in this path—FRT calculations show that frequency-noise-induced infidelity drops an order of magnitude upon cavity filtering and thus reducing laser frequency power spectral density.

It is also possible that increasing Rabi frequency and correspondingly increasing intermediate-state detuning will increase the light-shift caused by far-detuned atomic states. These light shifts will differentially affect $|1\rangle$ and $|r\rangle$, causing two-photon detuning to couple strongly to laser intensity. As Rabi frequency is increased, the gate will become increasingly sensitive to fluctuations in laser power. In the high-power limit, this source of error will dominate. It has been pointed out in the community [6] that this system is already in a regime of concern.

B. Use in Error Correcting Codes

The reported entangling gate fidelities are below the simulated threshold for the surface code. Using the methods described here, it is likely that systems of Rubidium

[6] Not naming names . . . Jeff.

atoms should be able to suppress noise with increasing code size. Suppressing infidelity to the three nines level will also have even more significant affect within the context of an error correcting code. To see this, consider the threshold equation:

$$p_L \propto \left(\frac{p}{p_{th}}\right)^{\lfloor (d-1)/2 \rfloor} \tag{11}$$

Given the extensive progress on quantum codes with a threshold (i.e., a p_{th} exists), it is imperative to furnish experimental progress in the direction of $p < p_{th}$ in order to take steps towards fault-tolerance. These series of works [1, 15] take that step for neutral atoms, passing the threshold for the surface code (which is 1%).

V. DISCUSSION QUESTIONS

1. Why do we model the dynamics of initial state $|11\rangle$ as being an effective two-level system with the $|W\rangle$ state? Why is their a negligible population of the state $|rr\rangle$ and precisely what approximation are we making here? Why are we not populating the orthonormal state $|V\rangle = (|1r\rangle - |r1\rangle)/\sqrt{2}$?

2. Seemingly, the objective for the control algorithm and the quantity we wish to estimate via Randomized Benchmarking look the same. What is the crucial difference?

3. Why are the fidelities in the 20-atom array case the same for the two gate protocols, given that the smooth amplitude suppresses intermediate scattering but the time optimal does not?

4. In figure (2c), why do they include X gates in their circuit? Why do they only use odd numbers of CZ gates?

5. What is intermediate-state scattering? How does the smooth amplitude gate reduce it? The smooth amplitude reduces scattering further if we increase gate times beyond what is done in the paper. Why don't they achieve extremely high-fidelities immediately by simply doing their gate more slowly?

VI. ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Prof. Jeff Thompson, particularly in the course QSE 503: Implementations of Quantum Information. The authors would also like to acknowledge Scout for moral support and inspiration.

[1] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **622**, 268–272 (2023).
[2] M. Saffman, T. G. Walker, and K. Mølmer, Rev. Mod. Phys. **82**, 2313 (2010).
[3] A. Browaeys and T. Lahaye, Nature Physics **16**, 132 (2020).
[4] G. Kleine Büning, J. Will, W. Ertmer, E. Rasel, J. Arlt, C. Klempt, F. Ramírez-Martínez, F. Piéchon, and P. Rosenbusch, Physical review letters **106**, 240801 (2011).
[5] M. Peper, Y. Li, D. Y. Knapp, M. Bileska, S. Ma, G. Liu, P. Peng, B. Zhang, S. P. Horvath, A. P. Burgers, and J. D. Thompson, "Spectroscopy and modeling of ^{171}yb rydberg states for high-fidelity two-qubit gates," (2024), arXiv:2406.01482 [physics.atom-ph].
[6] R. Finkelstein, R. B.-S. Tsai, X. Sun, P. Scholl, S. Direkci, T. Gefen, J. Choi, A. L. Shaw, and M. Endres, Nature **634**, 321 (2024).
[7] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Physical Review Letters **123** (2019), 10.1103/physrevlett.123.170503.
[8] D. Jaksch, J. I. Cirac, P. Zoller, S. L. Rolston, R. Côté, and M. D. Lukin, Phys. Rev. Lett. **85**, 2208 (2000).
[9] T. Wilk, A. Gaëtan, C. Evellin, J. Wolters, Y. Miroshnychenko, P. Grangier, and A. Browaeys, Phys. Rev. Lett. **104**, 010502 (2010).
[10] D. Gottesman, "An introduction to quantum error correction and fault-tolerant quantum computation," (2009), arXiv:0904.2557 [quant-ph].
[11] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).
[12] Nature **614**, 676 (2023).
[13] J.-S. Chen, E. Nielsen, M. Ebert, V. Inlek, K. Wright, V. Chaplin, A. Maksymov, E. Páez, A. Poudel, P. Maunz, and J. Gamble, Quantum **8**, 1516 (2024).
[14] M. Peper, Y. Li, D. Y. Knapp, M. Bileska, S. Ma, G. Liu, P. Peng, B. Zhang, S. P. Horvath, A. P. Burgers, and J. D. Thompson, "Spectroscopy and modeling of ^{171}yb rydberg states for high-fidelity two-qubit gates," (2024), arXiv:2406.01482 [physics.atom-ph].
[15] R. B.-S. Tsai, X. Sun, A. L. Shaw, R. Finkelstein, and M. Endres, "Benchmarking and fidelity response theory of high-fidelity rydberg entangling gates," (2024), arXiv:2407.20184 [quant-ph].
[16] S. Jandura and G. Pupillo, Quantum **6**, 712 (2022).
[17] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of magnetic resonance **172**, 296 (2005).
[18] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).
[19] J. Helsen, I. Roth, E. Onorati, A. Werner, and J. Eisert, PRX Quantum **3**, 020357 (2022).
[20] C. A. Sackett, D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. Turchette, W. M. Itano, D. J. Wineland, *et al.*, Nature **404**, 256 (2000).

[21] F. Roccati, G. M. Palma, F. Ciccarello, and F. Bagarello, Open Systems amp; Information Dynamics **29** (2022), 10.1142/s1230161222500044.

Appendix A: Scattering in the intermediate state

Consider a Markovian decay channel on a single qubit undergoing evolution under the Hamiltonian H, described by a prototypical GKSL master equation,

$$\dot{\rho} = -\iota[H,\rho] + \gamma\mathcal{D}[L]\rho \tag{A1}$$

where L is the jump operator and,

$$\mathcal{D}[L]\rho := L\rho L^\dagger - \frac{1}{2}(L^\dagger L\rho + \rho L^\dagger L) \tag{A2}$$

If we define,

$$H_{\text{eff}} := H - \gamma\frac{\iota}{2}L^\dagger L \tag{A3}$$

note that we can write,

$$\dot{\rho} = -\iota[H_{\text{eff}}\rho - \rho H_{\text{eff}}^\dagger] + \gamma L\rho L^\dagger \tag{A4}$$

Now, in attempt of a *semi-classical* treatment of this decay problem, one can neglect the 'jump' term and only deal with the non-Hermitian Hamiltonian evolution [21]. More formally, this problem is amenable to a treatment in the quantum trajectory formalism, and the non-Hermitian evolution (upon renormalization) is equivalent to post-selection on the no-jump trajectory. Taking $L = |\perp\rangle\langle e|$ to describe decay from the intermediate state, and $\gamma \equiv \gamma_e$, we see that $L^\dagger L = |e\rangle\langle e|$ and recover the form in the main text.

Appendix B: (A stab at) Bell state fidelity

The target Bell state is,

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{B1}$$

Now, the fidelity of a state ρ in the lab with the Bell state is,

$$F = \langle\Phi^+|\rho|\Phi^+\rangle = \frac{1}{2}(\langle 00|\rho|00\rangle + \langle 11|\rho|11\rangle + \langle 00|\rho|11\rangle + \langle 11|\rho|00\rangle) \tag{B2}$$

which is easily seen to be $= 1/2(P_{00} + P_{11}) + Q$ where P are populations and Q denotes coherence. This can in-turn be measured through the oscillation amplitudes of the two-atom parity operator $O := \sigma_1^z\sigma_2^z$ with a variable global $Z-$phase. More rigorously, this works as follows. Now, $\sigma^z = |0\rangle\langle 0| - |1\rangle\langle 1|$ and thus

$$\sigma_1^z\sigma_2^z = |00\rangle\langle 00| + |11\rangle\langle 11| - |01\rangle\langle 01| - |10\rangle\langle 10| \tag{B3}$$

Consider two states ρ and σ, with ρ being the state accessible in the lab and σ being the ideal state of the device. Let two POVM elements be $\Lambda_0 = |00\rangle\langle 00| + |11\rangle\langle 11|$ and $\Lambda_1 = |01\rangle\langle 01| + |10\rangle\langle 10|$ (adding up to one). The optimal measurement strategy is $\mathcal{M} = \{\Lambda_0, \Lambda_1\}$, and thus,

$$\frac{1}{2}||\rho - \sigma||_1 = \max_{0\le\Lambda\le 1} Tr(\Lambda(\rho - \sigma)) \le \epsilon \tag{B4}$$

where $||A||_1 := Tr(\sqrt{A^\dagger A})$ is the trace norm, intimately linked with the optimum measurement strategy to discriminate between two states. The idea is that, the Bell state is supposed to have $Tr(\Lambda_0\sigma) = 1$, so if we measure $Tr(\Lambda_0\rho) = 1-\epsilon$ in the experiment through the peak of the amplitude curves, we have $||\rho - \sigma||_1 \le 2\epsilon$ Now, we know that

$$||\rho - \sigma||_1 \le \epsilon \implies F(\rho,\sigma) \ge 1 - \epsilon \tag{B5}$$

Thus, we end up with a lower bound on the true fidelity.