Supplemental Information: A two-dimensional programmable tweezer array of fermions

MEASUREMENT OF FLOQUET HEATING

To measure the ground state lifetimes shown in Fig. 1 of the main text, we perform the following experiment. We load two tweezers, spaced 8.1 μm apart to avoid any overlap, each with a spin up and a spin down atom. Atoms in higher vibrational states are removed by lowering the tweezer depth and applying a magnetic field gradient, as described in [18]. The tweezer depth is increased to 50 kHz, and a variable hold time is applied. The tweezer light is strobed such that only one tweezer is on at a time, resulting in Floquet heating for low strobe frequencies. Finally, atoms in newly populated higher vibrational states are again removed with an identical spilling process, and the remaining $|\uparrow\rangle$ atoms are imaged.

THEORETICAL CALCULATIONS OF FLOQUET HEATING

We determine the Floquet heating from the strobing of the trap by directly calculating the dynamics of a single atom in a tweezer potential using a method based on a discrete variable representation (DVR) of Hilbert space [37], which was applied to ground state properties of tunnel-coupled tweezers in [14, 38]. The tweezer potential is

$$V(\vec{r}) = -\frac{V_0}{1+\frac{z^2}{z_R^2}} \exp\left[-\frac{2r^2}{w_0^2(1+\frac{z^2}{z_R^2})}\right] \tag{S1}$$

where V_0 is the trap depth when the tweezer is on, z_R is the Rayleigh range, and w_0 is the trap waist, and we use the measured parameters. We use a DVR basis of spatial-parity-adapted sinc functions on a three-dimensional cubic grid of (N_x, N_y, N_z) points spanning $(0,0,0)$ to (L_x, L_y, L_z) to represent our wavefunctions. Only even parity states occur for the dynamics considered.

The state at time t after the strobed time evolution is

$$|\psi(t)\rangle = \underbrace{\left[e^{-i\hat{T}\delta t}e^{-i(\hat{T}+\hat{V})\delta t}\right]\cdots\left[e^{-i\hat{T}\delta t}e^{-i(\hat{T}+\hat{V})\delta t}\right]}_{t/2\delta t \text{ times}}|i\rangle \tag{S2}$$

where $\hat{T} = -\hbar^2\nabla^2/2m + V_{\text{abs}}(\vec{r})$ is the kinetic energy operator ($\hbar$ is Planck's constant, and m is the mass of ^{6}Li) plus a potential $\hat{V}_{\text{abs}}$ to handle boundary conditions discussed below, $\hat{V}$ is the potential energy operator associated with Eq. (S1), and $\delta t = 1/(2f_s)$, and assuming $t/2\delta t$ is an integer (the micromotion between such times is discussed later). The initial state $|i\rangle$ is assumed to be the ground state of the system in the time-averaged potential, i.e. of $H_{1/2} = T + V/2$. Directly solving Eq. (S2) is challenging due to three factors: (1) the tweezer potential is 3D, (2) it involves a wide range of length scales (the wavefunction localization length, the Gaussian waist, and the Rayleigh range), and (3) the dynamics problem must account for a huge separation of timescales when f_s is large, the short time of the pulses and the long lifetime of the atoms, a ratio of timescales of more than 10^7.

We solve this by rewriting Eq. (S2). First we can diagonalize to obtain $e^{-i\hat{T}\delta t} = U_0 D_0 U_0^{-1}$ where D_0 is diagonal and U_0 is the matrix whose columns are eigenvectors of $\hat{T}$, obtained by first diagonalizing $\hat{T}$ and then exponentiating. (U_0 need not be unitary since $\hat{T}$ has non-Hermitian terms for the absorbing potential.) Similarly $e^{-i(\hat{T}+\hat{V})\delta t} = U_1 D_1 U_1^{-1}$,. Therefore the terms in brackets – denote it $M \equiv e^{-i\hat{T}\delta t}e^{-i(\hat{T}+\hat{V})\delta t}$ – in Eq. (S2) can be rewritten

$$M = U_0 D_0 U_0^{-1} U_1 D_1 U_1^{-1}. \tag{S3}$$

One can multiply these matrices, and diagonalize the result to find $M = U_2 D_2 U_2^{-1}$, so

$$|\psi(t)\rangle = M^{t/2\delta t}|i\rangle = U_2 D_2^{t/2\delta t} U_2^{-1}|i\rangle\,. \tag{S4}$$

The micromotion at times between integer multiples of $2\delta t$ can be treated by propagating to the largest multiple of $1/f_s$ as above, and then calculating the time evolution within a single period, which is straightforward and efficient in

terms of the diagonalized evolution matrices (D_a, U_a) already obtained. The computational cost is dominated by diagonalizing to find D_0, D_1, and D_2, with several additional multiplications. Although this involves full diagonalization of three large matrices, it avoids the issues associated with separation of timescales.

To obtain accurate results, one must use a a fine enough DVR grid and a large enough system. Here, a challenge arises that is absent for ground state calculations: atoms can be excited from the ground state to scattering states and thus escape from the trap. They may move rapidly, and thus would require L_a that grow linearly with time to capture, in practice many orders of magnitude larger than we can calculate. However, in the experiment once particles move sufficiently far outside the trap, the probability they return is negligible. This can be accurately reproduced in the calculations by including an absorbing boundary region in our simulations, a standard technique in calculations of, e.g., chemical dynamics [48]. We use a potential

$$V_{\rm abs}(\vec{r}) = -i\Gamma \sum_{a\in x,y,z} \frac{|a| - L_a^{(0)}}{L_a - L_a^{(0)}}. \tag{S5}$$

This is zero inside the cubic region from the origin to $(L_x^{(0)}, L_y^{(0)}, L_z^{(0)})$, and linearly increases as one moves outside this region. The whole calculation is performed inside a box of size (L_x, L_y, L_z) with (N_x, N_y, N_z) grid points. The calculations in the main text are performed with $(L_x^{(0)}, L_y^{(0)}, L_z^{(0)}) = (3, 3, 7.2)w_0$, $N = (27, 27, 23)$; $\Gamma = 57\,$kHz; and $L_a = L_a(0) + S_a$ where $S_a \approx w_0$, and are well-converged in the number of grid points and system size (Fig. S1). (S_a varies slightly in the right panel: because we fix the grid spacing and ensure that the last grid point in the non-absorbing region doesn't change with L_a, we choose S_a closest to w_0 while constraining L_a to correspond to the location of the last grid points along the a'th direction.) The Γ and S are chosen to be sufficiently large so that particles leaving the trap are absorbed before reaching the boundary of the calculation, but Γ is maintained sufficiently small so that there is no Zeno effect that would reflect the particles. The choices we make are consistent with these conditions, as given in Ref. [48], and our results are insensitive to the specific choice of the parameters. For example, varying Γ over a range of a factor of hundred has a negligible effect on the atom lifetime. Calculated lifetimes are converged to have visually negligible errors from the number of grid points, system size, and absorbing boundary.

Using this method, we calculate the population of the ground state (of the time-averaged potential) as a function of time. We find it gives a nearly perfect exponential decay for the f_s shown, with small deviations from exponential at smaller f_s. The lifetimes shown in Fig. 1(d) of the main text are obtained by fitting an exponential $e^{-t/\tau}$ to the obtained dynamics. The numerical calculations do not include the effects that give the finite lifetime in the unstrobed tweezer (background gas collisions and off-resonant light scattering), and which are expected to be the reason for the high-frequency lifetime saturation in Fig. 1(d). To incorporate those effects, the theory curve in Fig. 1(d) is a simple interpolation $1/\tau_{\rm eff} = 1/\tau + 1/\tau_{\rm static}$ where $\tau_{\rm static}$ is the lifetime in the unstrobed tweezer. The plot is largely insensitive to the details of this interpolation.

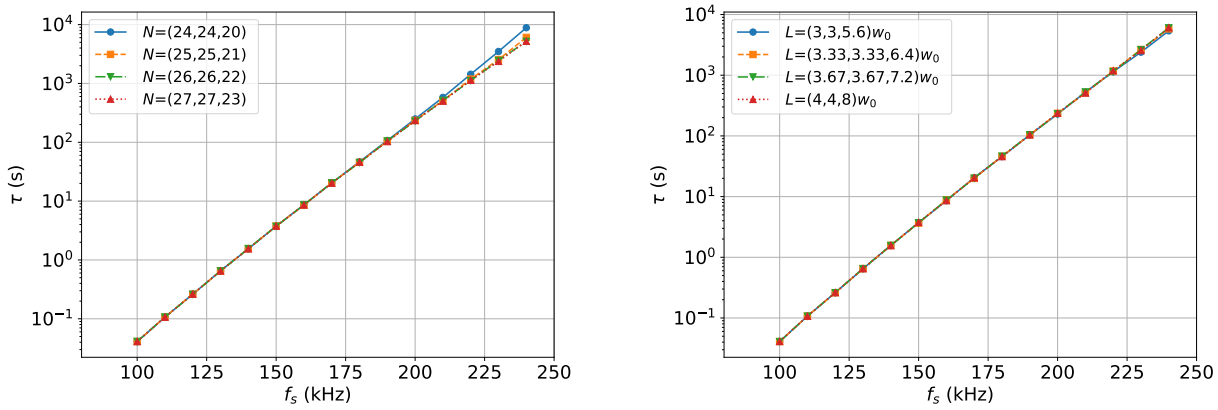

FIG. S1. Convergence of the calculated atom lifetime with respect to number of DVR grid points (left) and system size (right). Left: Lifetimes for different numbers of DVR grid points are converged well in the number of gridpoints for $(N_x, N_y, N_z) = (25, 25, 21)$ (or more) in systems of size $(L_x^{(0)}, L_y^{(0)}, L_z^{(0)}) = (3, 3, 7.2)w_0$. Right: lifetimes for different system sizes are converged well for $(L_x, L_y, L_z) = (3, 3, 5.6)w_0$ and larger. Right panel is at fixed grid point spacing $(0.16667, 0.166667, 0.4)w_0$, so the N_a vary with L_a. The spacing corresponds to $(N_x, N_y, N_z) = (24, 24, 20)$ for the largest system.

2D ARRAY GENERATION

Our arrays are generated by two perpendicularly oriented AOMs from AA Optoelectronic (MT110-B50A1,5-IR) driven by a two-channel arbitrary waveform generator (Spectrum M4i.6621-x8 PCIe). The typical strobing scheme, illustrated in Fig. 1 of the main text, generates one column of the array at any given time. This gives the most flexibility for creating different geometries. The two AOMs operate with the longitudinal acoustic mode with a high speed of sound (4200 m/s) in the crystal. This speed, along with the aperture size of $1.5 \times 2\,\mathrm{mm}^2$, sets the minimum possible dwell time on a tweezer (the duration for which the tweezer is on). We find that dwell times below $0.5\,\mu s$ produce tweezers with profiles that are significantly distorted along the strobe axis. When generating a two-site array by strobing, we measure that the tweezer waist along the strobe axis increases at 1.4% per increase of 100 kHz strobing rate. Distortions in the tweezer profile can be partly mitigated by applying a cosine-sum window function on the waveform that is sent to the strobed AOM. Given the minimum dwell time per tweezer, the strobe period $1/f_s$ grows linearly with the number of strobed columns. This seemingly limits us to about 8 sites along the strobe axis due to Floquet heating based on the results shown in Fig. 1 of the main text. However, we can surpass this limitation with the technique discussed in the following paragraph.

To reduce Floquet heating and distortion of the tweezer intensity profile, we combine 2D multitone arrays (produced by driving both AOMs with a set of radiofrequency tones) with strobing. For certain geometries, this allows us to reduce tweezer dwell times by using smaller modulation depths. We define modulation depth as the difference between the minimum and maximum relative intensity used over one strobe period. The principle is illustrated in Fig. S2 for 50% modulation depth, and was applied for the rectangular, Lieb, triangular, and ring lattices in Fig. 2 of the main text. For the example of Fig. S2, a static 2D multitone configuration (modulation depth of 0%) can be used in principle to create the geometry shown, but would not give enough degrees of freedom to homogenize the depths of all sites.

The type of lattices that can be generated with this approach is restricted to combinations of patterns that can be generated by using 2D multitone arrays. Assume each of the two AOMs is driven by a set of tones such that it individually produces tweezers at positions P_x, P_y respectively, where $P_x = \{p_{x,1}, ...p_{x,n}\}$ and $P_y = \{p_{y,1}, ...p_{y,m}\}$. The result when driving both AOMs will be the Cartesian product of the two sets $P_x \times P_y = \{(p_{x,i}, p_{y,j}) \,|\, p_{x,i} \in P_x, p_{y,j} \in P_y\}$. For example, triangular and Lieb lattices can be generated by switching between two multitone rectangular arrays and overlaying a small amplitude strobed array to homogenize the tweezer depths. In general, most 2D patterns of interest that host some level of periodicity can be generated by switching between two or three multitone rectangular arrays at full modulation depth and overlaying a small amplitude strobed array for homogenization. When using 2D multitone arrays, we avoid square geometries, as equidistant tweezer spacing in x and y leads to diagonal sites having degenerate tones, which results in low frequency beating between these tweezers and atom heating.

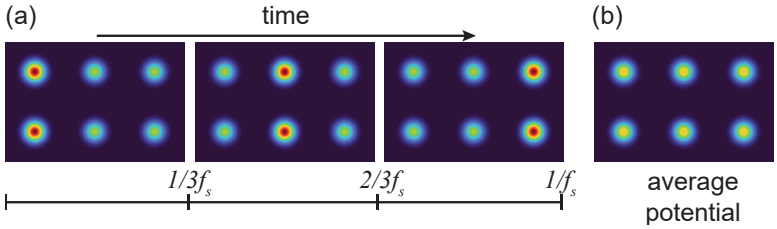

FIG. S2. Illustration of using 50% modulation depth to generate a 2×3 array with strobe rate f_s, with (a) the individual strobe steps and (b) the time-averaged potential. This procedure can be thought of as superimposing a static 2×3 grid on top of a 3-step array with 100% modulation depth.

2×2 PLAQUETTE

We calibrate the tunnelings in the horizontal and vertical directions in the 2×2 array as follows. We first load atoms on only one site, and reduce the tweezer intensities by 5 percent on all other sites except for the one in the direction we want to measure the tunneling. At a non-interacting magnetic field, we quickly initialize tunneling by zeroing the offsets among the target sites [18], and we measure Rabi oscillations between the two sites [14, 15].

To calibrate the on-site interaction energy U, we perform radiofrequency spectroscopy and measure the difference between single and double occupied sites. However, the most sensitive way to measure U/t is to fit the distribution of

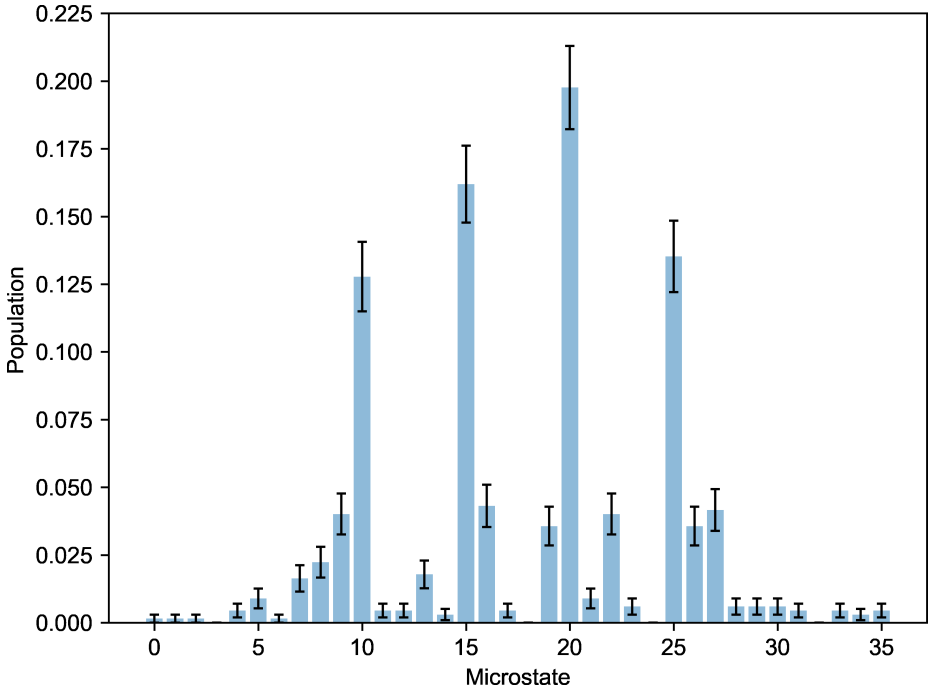

FIG. S3. Full microstate data for Fig. 4 of the main text without averaging to take advantage of the spin symmetry of the Hamiltonian.

microstates. In particular, for temperatures far below the interaction energy, U/t sensitively determines the number of single occupied sites. The value of U/t reported in the main text comes from a best fit of the experimental data, which is in reasonable agreement with an independent spectroscopic calibration [$U/t = 4.4(5)$].

To prepare the 2×2 plaquette ground state, we initialize the ground state of sites 1 and 4 with doublons, as pictured in Fig. 4 of the main text. To achieve a half-filled system, we turn on the opposing diagonal sites in a two-step sequence. First, the intensities of sites 2 and 3 are increased in 5 ms with an exponential time constant of 1 ms, until they are at ∼90% intensity of sites 1 and 4. Then, the intensities are further increased in a 50 ms exponential ramp with a time constant of 5 ms, until the offsets are zeroed. During this time, the magnetic field is ramped from a nearly non-interacting system at 573 G to its final value of 631 G using a two-point cubic spline function. A hold time of 4 ms is applied for further equilibration. Finally, tunneling is frozen by rapidly increasing the tweezer intensity to 2.5 times the science depth, and imaging commences.

In the main text of the paper, for the 2×2 tunneling data we average over states obtained by interchanging $|\uparrow\rangle$ and $|\downarrow\rangle$. This is because all potentials in the system are state independent. This reduces the statistical error due to the limited number of experimental shots (673 shots). Since we do not equalize the disorder entirely to zero, we do not assume any reflection symmetries that could be utilized in the ideal system to further reduce the statistical errors. We show the full microstate distribution without averaging in Fig. S3, and the full labelling of microstates in Tab. I.

ENTROPY AND TEMPERATURE

In order to fit the temperature of our system, we use the canonical ensemble as post-selection enables us to remove number fluctuations. Thus, with free parameters temperature T and U/t, the partition function is the sum over all energy eigenstates Ω, with

$$Z\left(U/t, T\right) = \sum_{\Omega} \exp\left(\frac{-E_{\Omega}(U/t)}{k_b T}\right). \tag{S6}$$

For each temperature and interaction, we create an expected distribution of microstates, and we compare to the experimental data using a weighted least squares fit. We find that the weighted sum of the squared residuals reaches

Index	$\|\uparrow\rangle$ population	$\|\downarrow\rangle$ population
0	1100	1100
1	1010	1100
2	1001	1100
3	0110	1100
4	0101	1100
5	0011	1100
6	1100	1010
7	1010	1010
8	1001	1010
9	0110	1010
10	0101	1010
11	0011	1010
12	1100	1001
13	1010	1001
14	1001	1001
15	0110	1001
16	0101	1001
17	0011	1001
18	1100	0110
19	1010	0110
20	1001	0110
21	0110	0110
22	0101	0110
23	0011	0110
24	1100	0101
25	1010	0101
26	1001	0101
27	0110	0101
28	0101	0101
29	0011	0101
30	1100	0011
31	1010	0011
32	1001	0011
33	0110	0011
34	0101	0011
35	0011	0011

TABLE I. The basis microstates for the data reported in Fig. 4 of the main text. The order of the sites used corresponds to the numbering in Fig. 4(a).

a local minimum (which depends very weakly on temperature) at $U/t = 3.4(2)$, where the errorbar is extracted from 500 bootstrapped samples of the data. We find that the temperature fit loses sensitivity below $k_{\mathrm{B}}T \sim 0.3\bar{t}$ (Fig S4). This corresponds to an entropy range of [0,0.09] k_{B} per particle. Exact diagonalization performed by in the Python package Quspin [49] suggests that for our ramp parameters, the entropy gain would be $0.02\,k_{\mathrm{B}}$ per particle.

In future studies, we wish to implement larger 2D ladder systems, preparing ground states of half-filled systems by the aforementioned technique of adiabatically ramping on additional tweezer sites. We use exact diagonalization on numerically tractable systems to estimate the many-body gaps sizes and therefore the feasibility of this approach. For instance, on a 2×5 triangular lattice at half filling, at $U/t = 6$, we can perform an adiabatic ramp in 100 ms that will only lead to an increase of 0.06 k_{B} per particle. For a 2×6 lattice with the same parameters, the increase is 0.07 k_{B}. These values were obtained by calculating dynamics with experimentally realistic ramp parameters. We are unable to perform exact diagonalization with our computing resources for larger 2D systems.

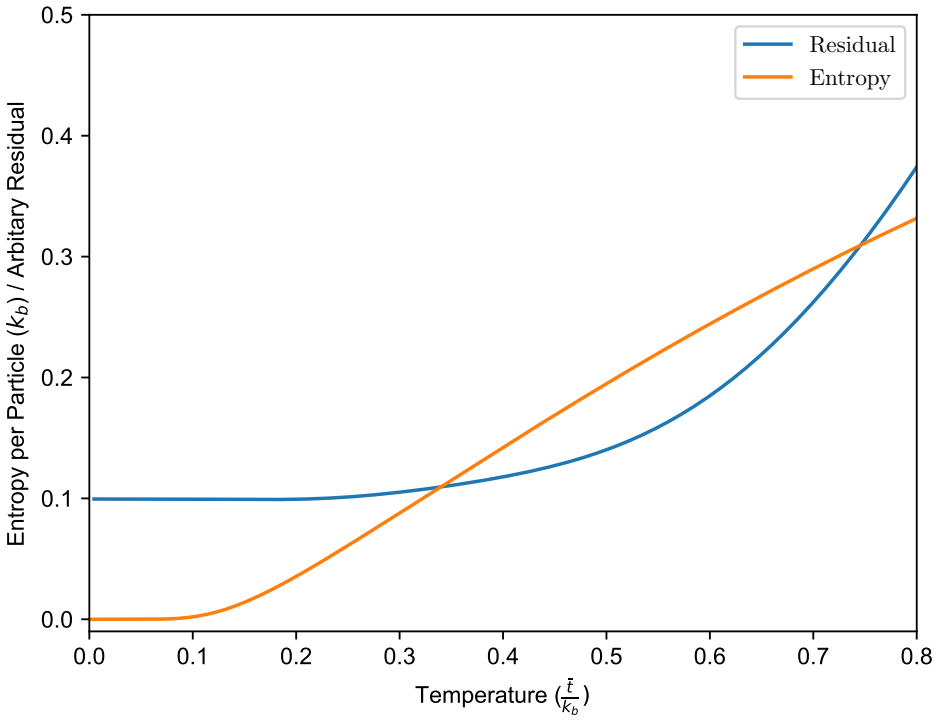

FIG. S4. Entropy and least squares residual as a function of temperature. Below around $k_{\mathrm{B}}T = 0.3\bar{t}$, the least squares residual is minimized, corresponding to an entropy range of [0,0.09] k_{B} per particle.