Quasimomentum Cooling and Relaxation in a Strongly Correlated Optical Lattice

David Chen,¹ Carolyn Meldgin,¹ Philip Russ,¹ Erich Mueller,² and Brian DeMarco¹,*

¹Department of Physics, University of Illinois
at Urbana-Champaign, Urbana, IL 61801, USA
²Department of Physics, Cornell University, Ithaca, NY 14853, USA

(Dated: March 31, 2015)
A key challenge for optical lattice experiments focused on simulating models of strongly correlated electronic solids has been achieving lower temperatures [1]. Although cooling of lattice-well vibrational states has been demonstrated [2–5], the motion of atoms through the lattice, characterized by their quasimomentum, has not been cooled directly. Furthermore, quasimomentum equilibration, which is necessary for cooling, has remained unexplored. Here, we use quasimomentum-selective stimulated Raman transitions to directly measure rethermalization rates and cool quasimomentum for a bosonic lattice Hubbard gas. The measured relaxation rate is consistent with a short-range, two-particle scattering model without free parameters at high lattice depths, despite an apparent violation of the Mott-Ioffe-Regel (MIR) bound [6]. Efficient cooling exceeding heating rates is achieved by iteratively removing high quasimomentum atoms from the lattice. Our results may have implications for models of unusual transport phenomena in materials with strong interactions, such as heavy fermion materials and transition metal oxides [6], and the cooling method we have developed is applicable to any species, including fermionic atoms.

Over the last decade, atoms trapped in optical lattices have proven to be a versatile platform for the study of Hubbard models [7], which are our simplest paradigms of strongly correlated electronic solids [8]. However, the inability to reach temperatures much lower than the magnetic super-exchange energy has limited access to new quantum regimes, such as the analog of $d$-wave superconductivity in the cuprates [1]. Several cooling schemes in optical lattices have been proposed, such as immersion cooling [9] and filter cooling [10–15], but experimental demonstrations have been limited [16, 17]. Notably, the kinetic energy, or quasimomentum degree of freedom has not been cooled directly and remains hotter than the Néel temperature in experiments with fermionic atoms [18].

We perform cooling and the relaxation measurements using quasimomentum-selective stimulated Raman transitions on a partially condensed Bose gas trapped in a cubic optical lattice. This technique uses a pair of Raman laser beams to transfer atoms from the ground band and the $|F = 1, m_F = 0\rangle$ hyperfine state to the first-excited band of the $|2, 0\rangle$ state (Fig. 1a). We label the hyperfine states as $|\downarrow\rangle$ and $|\uparrow\rangle$, respectively. Selectivity in quasimomentum is enabled by the difference in curvature of the ground and first-excited bands, and the targeted quasimomenta are controlled by a combination of the frequency difference between the Raman beams $\Delta\omega$, the effective Rabi rate, and the duration of the Raman pulse.
Because the wavevector difference $\Delta k = 1.45\pi/d$ ($d$ is the lattice spacing) lies along the $z$-direction of the lattice, the Raman transitions can only change the quasimomentum component $q_z$ by $\Delta q_z = 2q_B - \hbar\Delta k \approx 0.55q_B$ (Fig. 1b). The $|\uparrow\rangle$ atoms are ejected from the trap using light resonant with the $5S_{1/2}, F = 2 \rightarrow 5P_{3/2}$ transition.

We prepare a gas of $^{87}$Rb atoms in the $|\downarrow\rangle$ state with condensate fraction $N_0/N = 0.23 \pm 0.02$. The gas is confined in a 1064 nm dipole trap, and a cubic optical lattice formed from three pairs of counter-propagating laser beams is turned on in 100 ms to a lattice potential depth $s$ ranging from 4 to 8 $E_R$, where $E_R = \hbar^2\pi^2/2md^2$ is the recoil energy, $m$ is the atomic mass, $h = 2\pi\hbar$ is Planck’s constant, and $d \approx 406$ nm is the lattice spacing. Detailed information regarding our apparatus can be found in [19] and references therein. The temperature of the gas in the lattice is comparable with the bandwidth of the ground band. The atoms in the lattice are described by the Bose-Hubbard (BH) model with tunneling energy $t$ and interaction energy $U$, and the ratio $U/t$ is adjusted by tuning $s$ [20]. The range of $s$ and $U/t \approx 1$–7 accessed in these measurements corresponds to the superfluid regime of the BH model.

To measure rethermalization, the gas is brought out of equilibrium by rapidly removing atoms via pulsing on the Raman beams and resonant light consecutively for 200 $\mu$s and 50 $\mu$s, respectively, as shown in Fig. 1. The frequency difference of the Raman beams is tuned to remove the condensate and atoms with low quasimomentum $\vec{q}$ along the $z$ direction of the lattice. After the removal step, the remaining atoms are allowed to evolve in the lattice potential for a time $t_{\text{hold}} = (0 - 10)$ ms. The gas is then released from the lattice, and bandmapping and time-of-flight (TOF) imaging are used to measure the quasimomentum distribution [19, 21].

Characteristic TOF images used to measure the thermalization time $\tau$ are shown in Fig. 2. We fit each image to a semi-classical model that describes the equilibrium quasimomentum distribution of a non-interacting bosonic gas trapped in a combined lattice–parabolic potential. The mean squared residual of the fit $r^2 = \sum_{ij} (OD_{ij} - n_{ij})^2 / \sum_{ij} 1$ is used to quantify the deviation of the system from equilibrium, where $OD_{ij}$ is the measured optical depth, $n_{ij}$ is the fit function (see Ref. [19] and Supplementary Information), and the summations are over the indices $i$ and $j$ within a mask defined by the first Brillouin zone (FBZ). The residual $r^2$ is measured as $t_{\text{hold}}$ is varied at four different lattice depths: $s = 4, 5, 6$ and 8 $E_R$. 

3
FIG. 1. Quasimomentum-selective stimulated Raman transition and procedure for measuring thermalization. (a) Energy $\epsilon$ of $^{87}\text{Rb}$ atoms confined in a lattice potential. Only the ground band ($n = 0$) in $|\downarrow\rangle$ and first-excited band ($n = 1$) in $|\uparrow\rangle$ are shown. The Raman beams (red arrows) with Rabi rates $\Omega_1$ and $\Omega_2$ and detuned $\Delta$ from the transition to the $5P_{3/2}$ excited electronic state are used to selectively transfer atoms between bands. (b) Absorption images taken after bandmapping and 20 ms of TOF. The images are obtained before (top) and immediately after (middle) the Raman pulse and subsequent to the resonant-light pulse (bottom). The reciprocal lattice vectors are indicated by $k_x, k_y$, and $k_z$. The projection of the FBZ onto the imaging plane is displayed with dashed lines. Atoms in the first-excited band appear outside the FBZ after bandmapping.

We observe that the quasimomentum relaxation is exponential in time, as shown in Fig. 2 for the particular case of $s = 6 E_R$. The measured relaxation time constant $\tau$ determined from a simple exponential decay for the four lattice depths is shown in Fig. 3a. The relaxation

\[ 5P_{3/2} \]
\[ 5P_{1/2} \]
\[ \Delta \]
\[ 5S_{1/2} \]
\[ |\uparrow, n = 1\rangle \]
\[ |\downarrow, n = 0\rangle \]
process, which is approximately two orders of magnitude faster than that in the harmonic trap, speeds up for a stronger lattice potential, ultimately reaching just 200 µs at \( s = 8 \ E_R \). The relaxation is always faster than both Hubbard timescales, which vary from \( h/t = (3 - 9) \) ms and \( h/U = (2.6 - 1.3) \) ms for \( s = 4-8 \ E_R \).

We compare the measured time constant to a short-range, two-body scattering calculation based on Fermi’s golden rule (FGR), treating the Hubbard interaction term as a perturbation to the single-particle tight-binding Hamiltonian. This mechanism, which is analogous to electron-electron scattering, is distinct from the kinetic processes such as electron-phonon and electron-impurity interactions that dominate in most solid-state systems. We consider a thermal gas at equilibrium in a cubic lattice potential and calculate the relaxation rate of the occupation number \( n_{\vec{q}=0} \) to the equilibrium value after all the atoms at \( \vec{q} = 0 \) have been removed. The relaxation time \( \tau \) predicted by FGR is given by

\[
\frac{1}{\tau} = 4 \frac{\langle n \rangle}{R} F \left( \frac{t}{k_B T} \right) \frac{U^2}{t},
\]

where \( \langle n \rangle = N(m\omega^2 d^2/4\pi k_B T)^{3/2} \) is the filling-weighted average lattice filling, \( k_B \) is Boltzmann’s constant, \( N \) is the atom number, \( T \) is the temperature at equilibrium, and \( F(t/k_B T) \approx 0.3 \) (see Supplementary Information). Fig. 3b compares the prediction from Eq. (1) with the measured relaxation times. We show the normalized relaxation rate \( h/t\tau \) versus \( U^2/t^2 \) with \( \langle n \rangle \approx 0.13 \), which is the average value across all lattice depths determined by the lattice filling before relaxation. No free parameters were used for the theoretically predicted \( \tau \), which is constrained by the known experimental values.

We find excellent quantitative agreement between the predicted and measured relaxation rates at \( s = 6 \) and 8 \( E_R \). This agreement is surprising considering that the relaxation time is short compared with the time for an atom to tunnel between neighboring lattice sites, a condition that corresponds to a mean-free path shorter than the lattice spacing and a violation of the MIR bound [6]. The MIR limit is the requirement that the uncertainty in the quasimomentum of a particle must be less than the extent of the BZ for a semiclassical transport theory to be valid. The breakdown of the MIR bound has been associated with unusual transport phenomena in metallic compounds that are not completely understood. Remarkably, our simple kinetic theory works in this regime, even though it fails to capture the rethermalization time at lower lattice depths. The measured normalized relaxation time
FIG. 2. Sample rethermalization data. The mean squared residual of the quasimomentum-distribution fit $r^2$ versus hold time in the lattice $t_{\text{hold}}$ is shown for $s = 6 E_R$. Each data point corresponds to a single measurement, and the vertical axis of the plot has been re-scaled. The offset from $r^2 = 0$ is consistent with the residual from equilibrium images and is likely due to the failure of bandmapping at the edge of the FBZ and imaging noise. The red line is a fit to a single exponential decay used to determine the relaxation time constant $\tau$. The dashed white line shown superimposed on typical images taken at different hold times marks the FBZ.

does not scale linearly with $U^2/t^2$ and decreases less rapidly with $U^2/t^2$ than predicted by Eq. 1. The time constant $\tau$ is therefore smaller (and the rethermalization rate faster) at lower lattice depths than the FGR prediction. The failure of the scattering model at low
FIG. 3. Quasimomentum relaxation measurements at different lattice potential depths. (a) The measured relaxation time $\tau$ is shown for varied lattice potential depth $s$. Each point is determined using a fit to data such as those shown in Fig. 2, and the error bar displays the fit uncertainty. (b) Normalized relaxation rate $h/t\tau$ versus $U^2/t^2$. The dashed line represents the scaling law predicted by Eq. (1). The error bars include the fit uncertainty used to determine $\tau$ and the standard deviation in measurements of $T$ and $N$.

lattice depths may be related to the long-debated breakdown of quasiparticles and Fermi liquid theory in the low density limit of the Hubbard model (see Ref. [22], for example).

The extraordinarily fast quasimomentum relaxation we observe enables efficient cooling by iteratively removing the most energetic atoms of the gas from the ground band, in direct analogy to evaporative cooling [23]. In a proof-of-principle experiment, we have performed a series of two cooling cycles on a partially condensed gas trapped in a $s = 4 E_R$ lattice. The gas is initially prepared at the same temperature and atom number as for the relaxation experiment. For the cooling sequence (depicted in Fig. 4) the lattice is turned on in 100 ms, and the Raman beams and resonant light are subsequently pulsed on for 400 µs and 50 µs, respectively. Each Raman pulse is designed to excite only atoms with high quasimomentum, while the condensate is left largely unaffected. Given the spatial configuration of the Raman beams in our experiment, atoms are removed from one side of the FBZ in the $k_z$ direction, as shown in Fig. 4. At the end of a single cycle, a 1 ms delay is included to allow the
quasimomentum distribution to equilibrate. After all the cycles have completed, the gas is held in the lattice potential for 4 ms before bandmapping and imaging.

Figure 4a shows the quasimomentum profile of the gas along $k_z$ before cooling and after performing two cooling cycles. We observe that the width of the thermal component shrinks and the condensate number grows as the cooling sequence is performed. For a quantitative analysis, we fit the TOF images to the non-interacting, semi-classical model used in the relaxation analysis, with an additional independent Thomas-Fermi profile for the condensate. We use the fit to determine the number of thermal and condensate atoms (and the condensate fraction $N_0/N$) and $\beta t = t/k_B T$.

The results of these fits are shown in Figs 4b, c, and d. It is apparent in Fig. 4b that the thermal number decreases as atoms are expelled from the trap, while the condensate number increases. We interpret the redistribution of atoms from the thermal component to the condensate as a sign of rethermalization during the cooling cycle. Moreover, the steady increase in the condensate fraction evident in Fig. 4c indicates that the entropy per particle is reduced during cooling, and this technique may therefore be used to reach new quantum phases that exist at lower entropy. The decrease in temperature evident in Fig. 4d further suggests that the gas has achieved thermal equilibrium during cooling. Because bandmapping fails at the edge of the FBZ, a systematic error is made in the fitted $\beta t$. Nevertheless, the fitted $(\beta t)^{-1}$ is monotonically related to $T$ [19], and the decrease in $(\beta t)^{-1}$ signals a reduction in temperature.

A measure of the efficiency for any evaporative cooling scheme is $\alpha = d \log N / d \log T$. A smaller value of $\alpha$ indicates more efficient cooling—fewer atoms are removed for the same change in $T$. For our method, $\alpha = 1.75 \pm 0.04$ based on a fit to the data shown in Fig. 4. This performance compares favorably with recent results for non-lattice gases, including $\alpha \approx 1.5$ and 1.9 for “tilt” evaporation in a hybrid magnetic–optical trap [24, 25] and $\alpha \approx 2.7$ for dipole-trap evaporation of $^{87}$Rb [26].

The ultimate limit to the lowest temperature achievable by any cooling method is determined by competition between cooling and heating rates—cooling ceases when the two are equal [11]. The heating rate in optical lattices is primarily determined by momentum diffusion resulting from the interaction between the light and atoms [27]. This effect can be minimized by detuning the light far from any electronic transition. In our experiment, the heating rate
FIG. 4. Cooling of the quasimomentum distribution. The experimental sequence for cooling is shown schematically at the top; the timeline is not to scale. The TOF images show the quasimomentum distribution immediately before (i) and after (ii) atoms have been removed from the thermal component. (a) Quasimomentum profile along $k_z$ averaged over 3–4 images before (black) and after (gray-shaded) two cooling cycles. (b) Atom number in the condensate (▲) and in the thermal component (●) as the cooling cycles are performed. (c) and (d) show the condensate fraction and the fit parameter $(\beta t)^{-1}$ (which is monotonically related to the temperature) as a function of the cooling cycles. Plots (b)-(d) are obtained from fitting the TOF images to a semi-classical model. The error bars represent the standard error of the mean of the measurements averaged for each point.

induced by the lattice light is $0.15 E_R/s=25 \text{ pK}/\text{ms}$ at $s = 4 E_R$, while the cooling power
based on the data shown in Fig. 4d is approximately 9 nK $k_B/\text{ms}$ (corresponding to 0.6 $t/\text{ms}$). This extraordinary cooling power is possible because of the high thermalization rate. In the regime we explore, heating from the lattice is therefore not a limitation to the cooling method.

In principle, more cooling cycles could have been carried out in our experiment, since there are no technical limitations to this technique other than phase noise in the Raman pulses. However, drift in the initial temperature of the gas constrained our time window for data collection. The cooling efficiency can be improved by addressing both sides of the FBZ simultaneously using a small Raman wavevector $\Delta k$ compared with $\pi/d$ and by adding more pairs of Raman beams (see Supplementary Information). The ultimate cooling limit for this technique as applied to bosonic atoms is set by the selectivity in quasimomentum together with the finite momentum spread of the condensate. In the future, this method could be applied to fermionic atoms trapped in optical lattices in order to reach exotic quantum states that may exist at low entropy per particle. Further studies of relaxation in this system may contribute to our understanding of transport in materials that also display a violation of the MIR bound [6] and thermalization in closed quantum systems [28].

ACKNOWLEDGMENTS

E.M. acknowledges support from ARO grant W911NF-14-1-0003. B.D., D.C., C.M., and P.R. acknowledge support under grant PHY12-05548 and from the ARO under grant W9112-1-0462.

[1] D. C. McKay and B. DeMarco, Rep. Prog. Phys. 74, 054401 (2011).
[2] A. J. Kerman, V. Vuletić, C. Chin, and S. Chu, Phys. Rev. Lett. 84, 439 (2000).
[3] S. E. Hamann, D. L. Haycock, G. Klose, P. H. Pax, I. H. Deutsch, and P. S. Jessen, Phys. Rev. Lett. 80, 4149 (1998).
[4] H. Perrin, A. Kuhn, I. Bouchoule, and C. Salomon, EPL (Europhysics Letters) 42, 395 (1998).
[5] M. T. DePue, C. McCormick, S. L. Winoto, S. Oliver, and D. S. Weiss, Phys. Rev. Lett. 82, 2262 (1999).
[6] O. Gunnarsson, M. Calandra, and J. E. Han, Rev. Mod. Phys. 75, 1085 (2003).
[7] M. Lewenstein, A. Sanpera, and V. Ahufinger, Ultracold Atoms in Optical Lattices: Simulating quantum many-body systems (Oxford University Press, 2012).
[8] E. Dagotto, Rev. Mod. Phys. 66, 763 (1994).
[9] A. Griessner, A. J. Daley, S. R. Clark, D. Jaksch, and P. Zoller, N. J. Phys. 9, 44 (2007).
[10] F. Heidrich-Meisner, S. R. Manmana, M. Rigol, A. Muramatsu, A. E. Feiguin, and E. Dagotto, Phys. Rev. A 80, 041603 (2009).
[11] J.-S. Bernier, C. Kollath, A. Georges, L. De Leo, F. Gerbier, C. Salomon, and M. Köhl, Phys. Rev. A 79, 061601 (2009).
[12] T.-L. Ho and Q. Zhou, (2009), ArXiv:0911.5506.
[13] J. R. Williams, J. H. Huckans, R. W. Stites, E. L. Hazlett, and K. M. O’Hara, Phys. Rev. A 82, 011610 (2010).
[14] M. Popp, J.-J. Garcia-Ripoll, K. G. H. Vollbrecht, and J. I. Cirac, N. J. Phys. 8, 164 (2006).
[15] P. Rabl, A. J. Daley, P. O. Fedichev, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. 91, 110403 (2003).
[16] W. S. Bakr, P. M. Preiss, M. E. Tai, R. Ma, J. Simon, and M. Greiner, Nature 480, 500 (2011).
[17] P. Medley, D. Weld, H. Miyake, D. Pritchard, and W. Ketterle, Phys. Rev. Lett. 106, 195301 (2011).
[18] R. A. Hart, P. M. Duarte, T.-L. Yang, X. Liu, T. Paiva, E. Khatami, R. T. Scalettar, N. Trivedi, D. A. Huse, and R. G. Hulet, Nature 519, 211 (2015).
[19] D. McKay, M. White, and B. DeMarco, Phys. Rev. A 79, 063605 (2009).
[20] D. Jaksch, C. Bruder, J. Cirac, C. Gardiner, and P. Zoller, Phys. Rev. Lett. 81, 3108 (1998).
[21] J. H. Denschlag, J. E. Simsarian, H. Häffner, C. McKenzie, A. Browaeys, D. Cho, K. Helmersson, S. L. Rolston, and W. D. Phillips, J. Phys. B-At. Mol. Opt. 35, 3095 (2002).
[22] M. Y. Kagan, Modern Trends in Superconductivity and Superfluidity, Lecture Notes in Physics, Vol. 874 (Springer, 2013) Chap. 14.
[23] O. Luiten, M. Reynolds, and J. Walraven, Phys. Rev. A 53, 381 (1996).
[24] C.-L. Hung, X. Zhang, N. Gemelke, and C. Chin, Phys. Rev. A 78, 011604 (2008).
[25] Y.-J. Lin, A. R. Perry, R. L. Compton, I. B. Spielman, and J. V. Porto, Phys. Rev. A 79,
[26] A. J. Olson, R. J. Niffenegger, and Y. P. Chen, *Phys. Rev. A* **87**, 053613 (2013).

[27] J. Gordon and A. Ashkin, *Phys. Rev. A* **21**, 1606 (1980).

[28] M. Rigol, V. Dunjko, and M. Olshanii, *Nature* **452**, 854.

[29] S. Natu, D. McKay, B. DeMarco, and E. Mueller, *Phys. Rev. A* **85**, 061601 (2012).

[30] H. Metcalf and P. van der Straten, *Laser Cooling and Trapping*, Graduate Texts in Contemporary Physics (Springer New York, 1999).

[31] S. Koller, A. Groot, R. Duine, H. Stoff, and P. van der Straten, arXiv:1204.6143 (2012).

[32] B. DeMarco, S. Papp, and D. Jin, *Phys. Rev. Lett.* **86**, 5409 (2001).

[33] E. Timmermans, *Phys. Rev. Lett.* **87**, 240403 (2001).
SUPPLEMENTARY INFORMATION

Experimental parameters

The geometric mean of the trap frequencies is $\omega = 2\pi \times (55 \pm 5)$ Hz for the optical dipole trap. The atom number in the dipole trap before turning on the lattice is $(84 \pm 4) \times 10^3$ and the temperature is $(91 \pm 4)$ nK. The optical lattice formed from three pairs of counter-propagating laser beams with wavelength $\lambda = 812$ nm is turned on in 100 ms using an exponential ramp of the lattice light intensity. A 10.15 G magnetic field is applied to the atom gas to suppress spin-exchange collisions. The bandmapping procedure used for imaging transforms quasimomentum to momentum by ramping down the lattice quickly compared with $h/t$ but slowly with respect to $h/E_{bg}$, where $E_{bg}$ is the bandgap between the ground and first-excited bands. Quasimomentum redistribution during bandmapping can lead to a systematic error in condensate fraction [29]. For the range of condensate fraction explored here, that error is small and not dependent on condensate fraction.

For the quasimomentum relaxation measurement, the Raman beams are detuned by $\Delta = 2\pi \times 200$ GHz from the $D_1$ transition. The Raman beams are pulsed on for 200 $\mu$s, and the relative frequency is swept over 1.6 kHz across a central frequency from $\Delta \omega = \omega_1 - \omega_2 = \omega_0 + 2\pi \times (10 - 14.8)$ kHz, in correspondence to the lattice depths $s = (4 - 8) E_R$, where $\omega_0 = 2\pi \times (6834.742 \pm 0.001)$ MHz is the measured bare transition frequency between $|F = 1, m_F = 0\rangle$ and $|2, 0\rangle$ in a 10.15 G field. Different Rabi rates are used, ranging from $\Omega = 2\pi \times (3.2 - 4.2)$ kHz. For the cooling demonstration, the Raman beams are detuned by $\Delta = 2\pi \times 430$ GHz and they are pulsed on for 400 $\mu$s. The relative frequency ranges from $\Delta \omega = \omega_0 + 2\pi \times (15.8 - 16)$ kHz and the effective Rabi rate is $\Omega = 2\pi \times 1.0$ kHz. In both experiments, the resonant light is pulsed on for 50 $\mu$s.

A. Quasimomentum selectivity

In one dimension, the Raman excitation probability as function of quasimomentum is given by the formula

$$P(q_z) = \frac{\Omega^2}{\Omega^2 + \delta(q_z)^2} \sin^2 \left( \frac{\sqrt{\Omega^2 + \delta(q_z)^2} \Delta \tau}{2} \right)$$  \hspace{1cm} (S1)

where $\Delta \tau$ is the pulse duration, $\Omega = \Omega_1 \Omega_2 \langle \uparrow, n = 1 | e^{i\Delta k z} | \downarrow, n = 0 \rangle / 4\Delta$ is the effective Rabi
rate, $\Omega_{1,2}$ are the Rabi rates associated with the electric dipole transition $5S_{1/2} \rightarrow 5P_{1/2}$ for each Raman laser beam \cite{30}, $\delta(q_z) = \omega_1 - \omega_2 - [\omega_0 + (\epsilon^{(n=1)}(q_z + \Delta q_z) - \epsilon^{(n=0)}(q_z)) / \hbar]$ is the effective detuning, $\omega_{1,2}$ are the Raman laser frequencies, $\Delta q_z = 0.55q_B$, $\epsilon^{(n=0)}(q_z)$ and $\epsilon^{(n=1)}(q_z)$ are the tight-binding dispersion relations of the ground and first-excited bands respectively, and the frequency offset $\omega_0$ accounts for the hyperfine splitting and Zeeman shift between the $|\downarrow\rangle$ and $|\uparrow\rangle$ states. Fig. S1 shows $P(q_z)$ for the parameters used in the cooling experiment: $\Omega = (2\pi) \times 1$ kHz, $\Delta \omega = \omega_0 + (2\pi) \times 15.8$ kHz, and $\Delta \tau = 0.4$ ms.

FIG. S1. Raman excitation probability for the parameters used in the cooling experiment (red-solid). The ground and first-excited bands in the tight-binding approximation are also shown as a reference (black-solid). The hyperfine splitting is not shown. A smaller $\Delta q_z$ compared with $q_B = h\pi/d$ would permit both sides of the FBZ to be addressed simultaneously. As an example, the excitation probability for $\Delta q_z = 0.05q_B$ is shown as a dashed line. We ignore the weak dependence of $\Omega$ on $q_z$ for this calculation.
B. Image processing

The TOF images are fit to a semi-classical model that describes the equilibrium quasimomentum distribution of a non-interacting gas in a combined parabolic–lattice potential [19]. The quasimomentum distribution projected onto the imaging plane is given by

\[
n(q_1, q_2) = A \sum_{i=1}^{\infty} e^{-2i \beta (3 - \cos b)} \frac{z^i}{i^{3/2}} \left( (\pi - a) I_0(\alpha_i) + 2 \sum_{j=1}^{\infty} I_j(\alpha_i) \frac{\sin j(\pi - a)}{j} \right)
\]  

(S2)

where \(q_1\) and \(q_2\) are the horizontal and vertical directions in the absorption image, \(z\) is the fugacity, \(I_i\) is the modified Bessel function of the first kind, \(\alpha_j = 4j \beta t \cos a\), \(a = \pi |q_1 - q_1|/(2q_B)\), \(b = \pi (q_1 + q_2)/(\sqrt{2}q_B)\), \(\beta = 1/k_B T\), \(q_B = \hbar \pi /d\), \(d\) is the lattice spacing, and \(A\) is a constant. We perform the fit within a mask determined by the FBZ projected onto the imaging plane. A mask slightly smaller than the FBZ is used to account for the failure of bandmapping at the edge of the FBZ. For the rethermalization measurement, we mask out the residual condensate (corresponding to less than 6% condensate fraction) that remains in some images.

C. Relaxation rate calculation

To calculate the quasimomentum relaxation rate, we consider a Maxwell-Boltzmann thermal distribution in a uniform (i.e., without a trap) lattice potential with the \(\vec{q} = 0\) atoms removed. In the relaxation-time approximation, the Boltzmann equation predicts that the number of atoms \(n_0\) at \(\vec{q} = 0\) relaxes to the equilibrium value \(n_0^{eq} = e^{-\beta(\epsilon_0 - \mu)}\) exponentially, with a time constant \(\tau_{FGR}\) given by \(\partial_t n_0 = n_0^{eq}/\tau_{FGR}\). The time-rate-of-change \(\partial_t n_0\) is determined using Fermi’s golden rule:

\[
\partial_t n_0 = \frac{2U^2}{N^2} \sum_{pq} n_p n_q \frac{2\pi}{\hbar} \delta(\epsilon_{p+q} + \epsilon_0 - \epsilon_p - \epsilon_q),
\]  

(S3)

where \(\epsilon_p\) is the tight-binding dispersion relation for the ground band, \(n_p = e^{-\beta(\epsilon_p - \mu)}\), \(\beta = 1/(k_B T)\), \(k_B\) is Boltzmann’s constant, \(N\) is the atom number, and \(t\) and \(U\) are the Hubbard tunneling and interaction energies. The experimentally measured relaxation time \(\tau\) is given by

\[
\frac{1}{\tau} = \frac{2}{\tau_{FGR}} = 4 \frac{nF(\beta t)}{t} \frac{U^2}{t},
\]  

(S4)
where

\[
F(\beta t) = \frac{\int \frac{d^3\vec{p}}{(2\pi)^3} \frac{d^3\vec{q}}{(2\pi)^3} e^{2\beta t [C(\vec{p}) + C(\vec{q}) - 2C(0)]} 2\pi \delta (2C(\vec{p}) + 2C(\vec{q}) - 2C(\vec{p} + \vec{q}) - 2C(0))}{\int \frac{d^3\vec{p}}{(2\pi)^3} e^{2\beta t C(\vec{p}) - C(0))}},
\]

(S5)

\[C(\vec{q}) = \sum_{i=x,y,z} \cos(q_i),\]

and \(n\) is the lattice filling. Equation S4 includes a factor of 2 for comparison to the experiment, which measures a quantity proportional to the square of the density.

The parabolic confinement causes the lattice filling \(n\) to vary across the gas. We calculate the spatial average for \(n\) weighted by \(n\) assuming a Maxwell-Boltzmann density distribution, which results in \(\langle n \rangle = N(\beta m \omega^2 d^2 / 4\pi)^{3/2}\). The temperature of the gas in the lattice before removing atoms for the thermalization measurement ranges from \(T = (96 - 102)\) nK. The temperature is inferred using a thermodynamic calculation and matching entropy per particle \[19\]. To determine \(T\) before the Raman pulse, we use the entropy per particle determined from the condensate fraction measured in the dipole trap before turning on the lattice. As shown in Fig. S2, the function \(F(\beta t) \approx 0.3\) in the regime we measure the rethermalization rate (i.e., \(\beta t \approx 0.07 - 0.14\)).

D. Momentum relaxation in a harmonic trap

To test the method we use to measure the thermalization time in the lattice, we have measured momentum relaxation for a partially condensed gas trapped in the harmonic dipole potential. The gas is brought out of equilibrium by rapidly removing the condensate and atoms with low momentum along \(k_z\). The remaining atoms are allowed to relax for \(t_{\text{hold}} = (0 - 100)\) ms before performing TOF imaging. We fit each image to a gaussian and calculate the mean squared residual of each fit \(r^2\) (as defined in the main text).

We observe damped oscillations in \(r^2\) with a frequency matching that of the harmonic trap (Fig. S3). The data for \(t_{\text{hold}} > 15\) ms fit well to a damped sine-squared function. We exclude the data taken at short \(t_{\text{hold}}\) when the gas is far from equilibrium from the fit, since the relaxation in that regime does not appear to be exponential. The fitted damping time constant is \((40 \pm 10)\) ms, while the classical collision time is \(N/\Gamma = 150 \pm 40\) ms, where \(\Gamma = n_{\text{dwd}} v_{\text{rel}}\) is the total collision rate, \(N\) is the atom number, \(n_{\text{dwd}} = N^2 (m \omega^2 / 4\pi k_B T)^{3/2}\) is the density-weighted density, \(\sigma = 8\pi a_s^2\) is the elastic collision cross section, \(a_s\) is the \(s\)-wave scattering length \((a_s = 98a_0\) for \(^{87}\text{Rb}\), where \(a_0\) is the Bohr radius), and \(v_{\text{rel}} = (16k_B T / \pi m)^{1/2}\)
FIG. S2. Function $F(\beta t)$ from the FGR calculation.

is the mean relative speed.

To compare the time constant to the collision time, we determine the average number of collisions required for rethermalization using a 3D molecular dynamics simulation of hard spheres confined in a harmonic trap. We simulate 5000 particles with an initial Maxwell-Boltzmann distribution of velocities and positions. At the beginning of the simulation, the particles with speeds 20% smaller than the average are removed along one direction. Particle trajectories are then propagated using the velocity-Verlet method. When an overlap between two particles is detected, the direction of the relative velocity is randomized (while preserving the center-of-mass velocity of the pair) to simulate $s$-wave collisions. The velocity distribution is analyzed at each time step to compute $r^2$ from the equilibrium distribution. We observe behavior similar to that displayed in Fig. S3. A fit of $r^2$ vs. the evolution time to an exponentially damped oscillatory function indicates that the time constant corresponds to 0.5 collision per particle for a wide range of particle diameters.

We conclude that the measured time constant for relaxation in the dipole trap is consistent
FIG. S3. Mean squared residual $r^2$ vs. holdtime $t_{\text{hold}}$ for a thermal gas in a harmonic trap. The vertical axis has been re-scaled. The linear regime is fit to a damped sine-squared function (solid red line).

with $N/\Gamma$ within a factor of two. The discrepancy between the measured time constant and the inferred collision rate may arise from systematic uncertainty in the number of particles, trap anharmonicity, and quantum degeneracy effects [31].

E. Simulation of cooling quasimomentum

We perform a simple simulation of the potential enhancement in cooling efficiency when more pairs of Raman beams are added to simultaneously remove thermal atoms along one, two, and three lattice vector directions. For the simulation, the cooling sequence is performed iteratively on a 3D quasimomentum distribution based on Eq. [S2]. The atom-removal step
is modeled via truncation of the quasimomentum distribution beyond a threshold, and the rethermalization step is modeled via calculation of the new equilibrium temperature of the system through energy conservation. A heating rate associated with the laser beams (for $s = 4 E_R$) is also included in the simulation.

In Fig. S4 we show the entropy-per-particle of the gas $S/N$ versus the number of iterations of the cooling cycle. The 1/2D line (black) corresponds to the case where atoms are removed from one side of the FBZ at a time along one of the lattice vectors, as we have demonstrated experimentally. The 1D line (red) represents the case where atoms on both sides of the FBZ are addressed simultaneously. This can be achieved by using a single pair of Raman beams with a small wavevector $\Delta k$, as shown in Fig. S1. The 2D (blue) and 3D (purple) cases refer to the extension of the technique into the orthogonal directions using one and two additional pairs of beams. Fig. S4 shows that the cooling efficiency improves as the technique is extended beyond 1D, and that $S/N$ converges to zero in a reasonable number of steps.

We conclude that there is no fundamental limitation to cooling via this method for weakly interacting systems. Furthermore, realistic heating rates from the lattice laser beams are not a complication. Ultimately, Pauli blocking [32] and hole heating [33] may limit the lowest temperatures achievable by this method for fermionic atoms. Interaction effects, which will require a more sophisticated analysis, may also introduce difficulties.
FIG. S4. Simulation of the entropy-per-particle versus number of cooling iterations in 1/2D (black, solid), 1D (red, dotted), 2D (blue, dashed) and 3D (purple, dotted-dashed). The initial atom number is $N = 1 \times 10^5$ and the initial temperature is $T = 50$ nK. The initial quasimomentum cutoff is $0.9 q_B$ and the final cutoff is $0.02 q_B$. The estimated heating rate is $0.15 E_R/s$. 