

Supplementary Materials for Direct Measurement of Room Temperature Non-diffusive Thermal Transport Over Micron Distances in a Silicon Membrane

Si Membrane Fabrication

Freestanding Si membranes were fabricated in a nominally undoped, silicon-on-insulator (SOI) wafer using Si MEMS processing techniques (see Fig. S1). In this process, the underlying Si substrate and buried oxide layer are removed through a combination of dry and wet etching techniques to leave a top layer of suspended silicon.

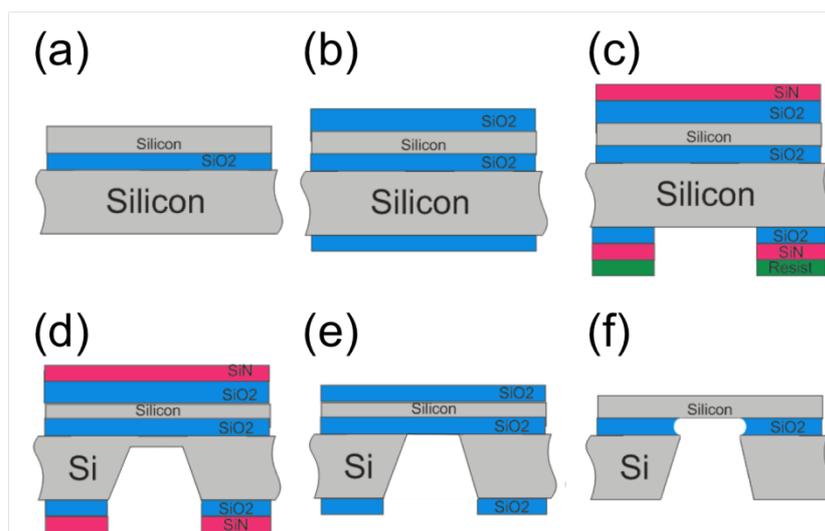


FIG. S1. Process flow for the fabrication of freestanding Si membranes: (a) Original SOI wafer. (b) Oxidation process to reduce thickness of Si layer. (c) Deposition of Si_3N_4 , photolithography of the etching window, including spin-coating, exposure, and development of the resist layer, and RIE of the Si_3N_4 and SiO_2 layers. (d) Wet etching of the Si substrate using KOH, until approximately $1\ \mu\text{m}$ of Si remains. (e) Finishing of the etching of the substrate with TMAH, followed by removal of the Si_3N_4 by RIE. (f) Final wet etching using HF to remove the top protective oxide and the buried oxide layers, releasing the freestanding Si membrane.

The initial SOI wafer was $625\ \mu\text{m}$ thick in total, with a top Si layer thickness of $\sim 1.5\ \mu\text{m}$ and a $\sim 1\ \mu\text{m}$ buried oxide layer. The target thickness of the top Si layer was achieved through oxidation of the wafer; as the thermal oxide incorporates silicon during growth, the thickness of the top Si layer is reduced. For every unit thickness of Si consumed, 2.27 unit thickness of oxide is grown. The oxidation was performed at a temperature of $1100\ ^\circ\text{C}$ in an atmosphere of water vapor in two steps to achieve fine control over the growth process, due to the large initial thickness of the top Si layer. Oxidation was continued until approximately $400\ \text{nm}$ of Si remained on the top layer, and the thermal oxide was left as a protective layer, as illustrated in Fig. S1(b), until the final stage of the process.

A silicon nitride layer was deposited to act as a mask during the subsequent wet etching of the Si substrate and the freestanding areas of the membranes were determined through photolithography on the backside of the wafer, involving spin-coating of a photoresist, exposure, and development. The remaining photoresist was then used as a mask for Reactive Ion Etching (RIE) to open etching windows in the Si_3N_4 and SiO_2 (see Fig. S1(c)).

A wet etching process with potassium hydroxide (KOH) and tetramethylammonium hydroxide (TMAH) was used to remove the Si substrate. The selectivity of TMAH is better when using a Si₃N₄ or SiO₂ mask for the etching of Si, though the etch rate is slower. The layer of Si₃N₄ was deposited to improve the etch selectivity compared to the SiO₂, and the Si substrate was etched with KOH until approximately 1 μm of Si substrate remained (Fig. S1(d)). The etching of the substrate was finished with TMAH, and the Si₃N₄ was then removed by RIE. (Fig. S1(e)). The etching occurs preferentially in the <100> direction with an etching angle of 54.7 degrees, thus the areas of the membranes are substantially smaller than the original patterns on the backside of the wafer. The reduction in length of one side can be calculated approximately by $x_f = x_i - 2d_{sub}$, where d_{sub} is the thickness of the Si substrate, x_i is the initial length on the backside of the wafer, and x_f is the final length on the topside of the wafer. This relaxes the resolution requirements for the photolithography, and allows the photolithography masks to be produced from inexpensive, disposable acetate in place of quartz.

After the wet etching process of the Si substrate, the top, bottom, and buried oxides were removed by a wet etch of hydrofluoric acid (HF) to release the freestanding Si membranes (Fig. S1(f)). Measurements were conducted on two membranes with 400×400 μm² freestanding area and thicknesses, determined by optical reflectometry, of 400±10 nm. Both membranes looked identical by visual inspection under a microscope.

Experimental Details

A short-pulsed excitation laser beam was derived through the frequency doubled output ($\lambda_e = 515$ nm) of an amplified Yb:KGW laser system (HighQ femtoRegen, set to 1 kHz repetition-rate). Although the laser is designed to output pulses as short as 300 fs in duration, to avoid sample damage and unwanted nonlinear optical effects from high peak powers, we have bypassed the compressor to obtain ~60 ps pulses. As depicted in Fig. S2, the pump beam was split with a custom diffractive optic (a binary phase mask pattern) into two beams which were passed through a two-lens telescope (with 2:1 imaging by achromatic doublets) and were focused and crossed within the membrane, with 3.6 μJ per pulse and the spot size radius 300 μm at 1/e intensity level. The CW probe beam, derived from the output of a single-longitudinal-mode, intracavity frequency-doubled Nd:YAG laser at 532 nm, followed an almost identical optical path as depicted in Fig. S2 and was focused into a spot of 150 μm 1/e radius aligned with the center of the excitation spot. To reduce sample heating, an electro-optic modulator was used in conjunction with a delay generator to chop the probe beam into 64 μs rectangular pulses synchronized with the pump pulses. An absorptive neutral density filter was used to attenuate the 5.2 mW reference beam power by a factor of 1000 to avoid detector saturation. The heterodyne phase was controlled by small angle adjustments of a highly parallel fused silica plate placed in the probe beam path. The co-propagating reference and diffracted probe beams were directed to a Hamamatsu C5658 silicon avalanche photodiode with 1 GHz bandwidth and the signal traces were recorded on an oscilloscope with 4 GHz bandwidth. Traces of 40,000 averages were subsequently downloaded to the computer for data analysis. The measurements were conducted in air and the temperature in the lab varied between 296 and 298 K.

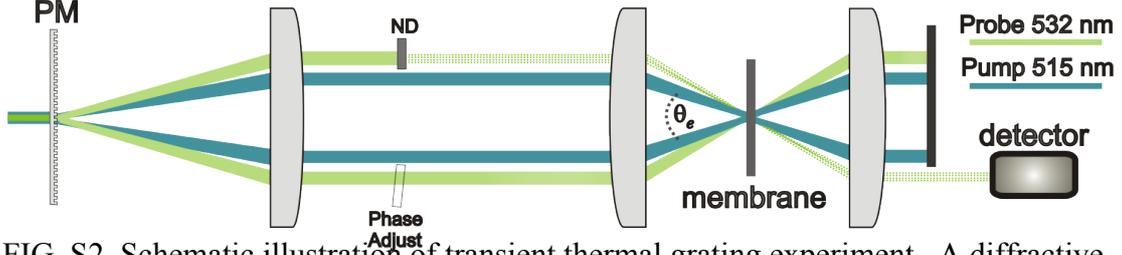


FIG. S2. Schematic illustration of transient thermal grating experiment. A diffractive optic, a binary phase-mask (PM), splits pump and probe into ± 1 diffraction orders. Pump beams are crossed in the silicon membrane, generating the transient thermal grating. Diffracted probe light is combined with a reference beam attenuated by a neutral density filter (ND) and directed to a fast detector. The relative phase difference between probe and reference beams is controlled by adjusting the angle of a glass slide (Phase Adjust) in the probe beam path.

The Probing Process and Heterodyne Detection

The excitation processes and resulting material responses described above lead to time-dependent, spatially periodic changes in both the complex refractive index of silicon and the thickness of the membrane; the dynamics of these induced transient grating responses will be encoded in the diffracted probe light that is directed to a detector. The optical fields of the probe and reference beams incident on the sample are approximated, respectively, as plane waves

$$E_p = E_{0p} \exp\left(i\left(k_p^2 - q^2/4\right)^{1/2} z - i(q/2)x - i\omega_p t + i\phi_p\right) \quad (\text{S1})$$

and

$$E_R = t_r E_{0p} \exp\left(i\left(k_p^2 - q^2/4\right)^{1/2} z + i(q/2)x - i\omega_p t + i\phi_R\right) \quad (\text{S2})$$

where E_{0p} is the incident probe amplitude, t_r is the attenuation factor for the reference beam, q is the transient grating wavevector, k_p is the optical wavevector magnitude, ω_p is the optical frequency, and ϕ_p and ϕ_R are the phases of probe and reference beams respectively.

The diffracted field can be obtained by multiplying the input field by the complex transfer function (S1) for an absorbing slab (S2). The transfer function depends on the complex refractive index and the thickness, both of which are functions of temperature. Assuming the temperature grating to be a small perturbation, the transfer function can be represented as

$$t = t_0 [1 + a(T_u + T_g \cos(qx))], \quad (\text{S3})$$

where the temperature perturbation is comprised of a spatially uniform component T_u and the grating component T_g , t_0 is the transmission of the slab in the absence of temperature perturbations, and a is the magnitude of the temperature-induced transmission change taking into account changes in the complex refractive index \tilde{n} and the sample thickness d ,

$$a = \frac{1}{t_0} \left(\frac{\partial t}{\partial \tilde{n}} \frac{\partial \tilde{n}}{\partial T} + \frac{\partial t}{\partial d} \alpha d \right), \quad (\text{S4})$$

where α is the thermal expansion coefficient. Derivatives $(\partial t / \partial \tilde{n})$ and $(\partial t / \partial d)$ can be obtained from the known expression for the transmission amplitude of a slab (S2).

For example, for normal incidence we get

$$\frac{1}{t_0} \frac{\partial t}{\partial \tilde{n}} = \frac{2}{\tilde{n}} \frac{\tilde{n} + 1 - (\tilde{n} - 1) \exp(2ik_p \tilde{n}d)}{(\tilde{n} + 1)^2 - (\tilde{n} - 1)^2 \exp(2ik_p \tilde{n}d)} \quad (\text{S5})$$

$$\frac{1}{t_0} \frac{\partial t}{\partial h} = ik_p h \frac{(\tilde{n} + 1)^2 + (\tilde{n} - 1)^2 \exp(2ik_p \tilde{n}d)}{(\tilde{n} + 1)^2 - (\tilde{n} - 1)^2 \exp(2ik_p \tilde{n}d)}$$

In our case, the incidence angle of the probe beam on the sample does not exceed 6° , therefore Eq. (S5) yields a very reasonable approximation. At any rate, knowing the value of amplitude factor a is not essential because our study is based on measuring the temporal dynamics of the signal rather than its magnitude.

Assuming that the sample is located at $z = 0$, for the +1 diffraction order of the probe beam one obtains

$$E_{p(+1)} = \frac{1}{2} t_0 a T_g E_{0p} \exp\left(i(k_p^2 - q^2/4)^{1/2} z + i(q/2)x - i\omega_p t + i\phi_p\right) \quad (\text{S6})$$

and for the zero order reference beam

$$E_{R(0)} = t_r t_0 (1 + a T_u) E_{0p} \exp\left(i(k_p^2 - q^2/4)^{1/2} z + i(q/2)x - i\omega_p t + i\phi_R\right) \quad (\text{S7})$$

The two beams are collinear and their interference gives an intensity

$$I_s = \frac{1}{2} I_{0p} |t_0|^2 \left[t_r^2 |1 + a T_u|^2 + |a|^2 T_g^2 + t_r T_g (a e^{i\phi} + a^* e^{-i\phi}) + 2 t_r |a|^2 T_u T_g \cos \phi \right], \quad (\text{S8})$$

where I_{0p} is the intensity of the probe beam, a^* is the complex conjugate of a , and $\phi = \phi_p - \phi_R$ is the heterodyne phase. The heterodyne phase is well defined and easily adjusted by rotating the thin glass plate in the probe beam path (Fig. S2). Neglecting terms quadratic with respect to the temperature perturbation, we get

$$I_s = \frac{1}{2} I_{0p} |t_0|^2 \left[t_r^2 (1 + 2 T_u \text{Re } a) + 2 t_r T_g (\text{Re } a \cos \phi + \text{Im } a \sin \phi) \right]. \quad (\text{S9})$$

We collect the signal for two values of the heterodyne phase ϕ separated by π and calculate the difference equal to

$$I_{diff} = t_r |t_0|^2 I_{0p} T_g (\text{Re } a \cos \phi + \text{Im } a \sin \phi). \quad (\text{S10})$$

The dynamics of this difference signal are entirely determined by the amplitude of the temperature grating T_g . The heterodyne phase is adjusted to maximize the signal; the exact value of the phase does not affect the signal shape. Taking the difference of the signals measured at two opposite heterodyne phases also helps suppress spurious signals, such as Pockels' cell interference, that are not sensitive to the heterodyne phase.

Besides the temperature grating, laser excitation also initially creates a grating of the photoexcited carriers concentration which modulates both the refractive index and the membrane thickness (the latter changes due to strain induced via deformation potential [S3]). Including this effect in the analysis is straightforward: the perturbed transfer function $t = t_0 [1 + aT + bN]$ now depends on both the temperature perturbation and photoexcited carrier concentration N , comprised, again, of a uniform and “grating” components, $N = N_u + N_g \cos(qx)$. Repeating the steps leading to Eq.(S10) we get the following result, reflecting the fact that in the small perturbation case thermal and electronic effects must be additive,

$$I_{diff} = t_r |t_0|^2 I_{0p} \left[T_g (\text{Re } a \cos \phi + \text{Im } a \sin \phi) + [N_g (\text{Re } b \cos \phi + \text{Im } b \sin \phi)] \right]. \quad (\text{S11})$$

If the decay of T_g and N_g occurred on the same time scale, isolating the temperature dynamics would be very difficult. However, as will be shown below, the carrier grating decays much faster because the ambipolar carrier diffusion coefficient in Si is much larger than the thermal diffusion coefficient. After the photoexcited carrier grating is washed away by carrier diffusion and relaxation, we are left with a purely thermal grating whereby the signal is described by Eq. (S10).

Electronic Decay and Data Analysis

Even though the carrier dynamics in Si at these length and time scales have been well studied and are not the focus of the current study, a basic understanding of the electronic response is important for two reasons: 1) as mentioned above, if the decay of T_g and N_g occur on the same time scale, isolating the temperature dynamics would be difficult, and 2) could the excited carrier population in some way influence the phonon MFPs to account for the effects we observe? The electronic response is expected to follow a diffusion equation similar to the thermal response, which is given by [S4]:

$$\frac{dN(x,t)}{dt} = -N/\tau - BN^2 - \gamma N^3 + D_a \frac{\partial^2}{\partial x^2} N \quad (\text{S12})$$

where $N(x,t)$ is the carrier concentration, τ is the band-edge carrier lifetime through single carrier decay channels, B is the radiative recombination coefficient, γ is the Auger recombination coefficient, and D_a is the ambipolar diffusivity. The diffusion term gives the wavevector dependence of the electronic decay. Due to the ambipolar diffusion coefficient in Si being much larger than the thermal diffusivity ($5 \text{ cm}^2/\text{s}$ or greater for carriers, depending on the carrier concentration [S4], and $0.86 \text{ cm}^2/\text{s}$ for diffusive thermal transport in bulk Si at 300K [S5]), the electronic and thermal transient grating signal decay components will always be well separated in time. The fact that carrier relaxation through various channels also affects the carrier concentration dynamics alters the expected diffusive q^2 decay, particularly for low q , to help speed the electronic decay, making the separation between electronic and thermal decays only larger, and our isolation of thermal dynamics easier. An earlier transient grating study on bulk Si [S6] also found that carrier diffusion occurs much faster than thermal diffusion and that electronic and thermal relaxation dynamics are well separated.

To measure the decay rate of the thermal grating, one could exclude the initial fast electronic response and fit the rest of the signal to a single exponential decay. However, this could introduce a small systematic error depending on the position of the fit starting point. Thus we chose to account for the electronic response directly by fitting the signal to a bi-exponential form as illustrated in Fig. S3(a) for the signal waveform obtained at $L=10 \text{ }\mu\text{m}$. Single-exponential components of the fit function corresponding to the electronic and thermal components of the signal are shown as well; one can see that the two components are well separated in time. Figure S3(b) shows good agreement between the data and bi-exponential fit for four representative grating periods (note the logarithmic time scale chosen to show both electronic and thermal dynamics at different grating periods). Oscillations present in the $2.4 \text{ }\mu\text{m}$ signal are due to the detector response to the fast initial electronic signal.

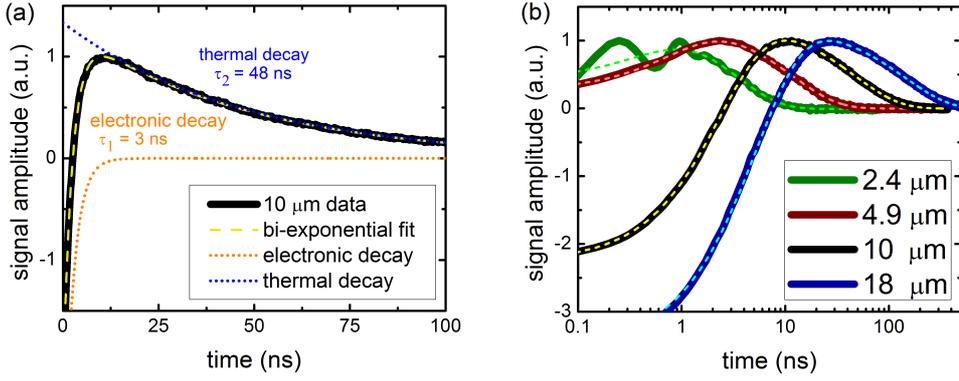


FIG. S3. (a) Data trace for the 10 μm transient grating period showing the bi-exponential fit (dashed line) in tandem with fast electronic and slower thermal exponential decays (dotted lines). (b) Full traces of 2.4, 4.9, 10, and 18 μm transient grating periods with accompanying bi-exponential fits (dashed lines).

In Figure S4, we see the measured electronic decay, extracted from the bi-exponential fit as a function of q^2 . We clearly see that electronic and thermal decay rates are significantly different for all measured wavevectors.

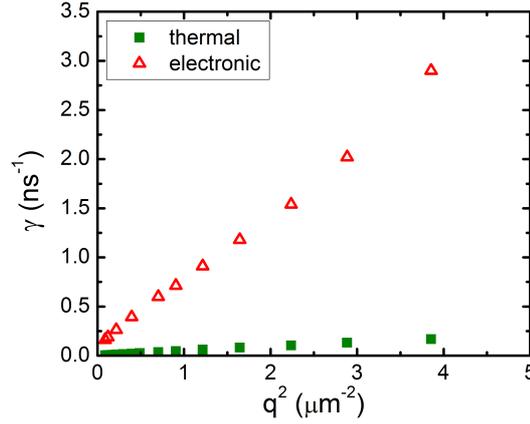


FIG. S4. Measured electronic (open triangles) and thermal (filled squares) decay rates vs q^2 . We see that for all wavevectors the electronic and thermal decay rates are well separated, allowing unambiguous determination of the thermal decay rate.

We have shown that in the transient grating signal, sensitive to the periodic carrier and temperature distributions, electronic and thermal responses are well separated at all grating periods, making unambiguous determination of the thermal dynamics possible. In indirect-band gap silicon, even though the electronic diffusion is fast, complete carrier relaxation still may take some time. For the carrier concentrations relevant to the current measurements, Auger recombination is the dominant electronic decay mechanism, and therefore, according to Eq. (S12), the electronic dynamics depend strongly on the initial carrier concentration. We measured the complete relaxation dynamics as a function of pump pulse energy, which varies the initial carrier concentration. For the different pump energies, with initial carrier concentrations varying according to our estimates from 5×10^{18} to $2 \times 10^{19} \text{ cm}^{-3}$, revealed no change in the thermal dynamics; the thermal decay rates were identical, indicating that excited carriers do not interact strongly with heat carrying phonons to

significantly reduce the MFPs and suggesting that there is no connection between our observations and the excited carrier concentration.

Thermal Grating Relaxation in a Thin Membrane in Ambient Medium

We consider thermal grating relaxation in a thin membrane of thickness $2h$ in an ambient medium (i.e., air). Due to symmetry, this problem is equivalent to that of a film of thickness h on a substrate occupying a half-space $z > 0$. Heat transport in the substrate is described by the heat diffusion equation

$$\frac{\partial T}{\partial t} = \alpha_s \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2} \right) \quad (\text{S13})$$

where α_s is the thermal diffusivity of the substrate. For a thermally thin film, the film temperature is assumed to be uniform over the film thickness and equal to the substrate temperature at $z = 0$. The heat transport equation for the film incorporating heat exchange with the substrate and the laser heating source $Q(t,x)$ can then be written as

$$\sigma \frac{\partial T}{\partial t} = \sigma \alpha_f \frac{\partial^2 T}{\partial x^2} + \lambda_s \frac{\partial T}{\partial z} + Q(t,x), \quad z = 0 \quad (\text{S14})$$

where σ is the specific heat per unit area of the film, $\sigma = \rho_f C_f h$, with ρ_f and C_f being the density and the specific heat of the film material, and $\lambda_s = \rho_s C_s \alpha_s$ is the substrate thermal conductivity. Let us model the laser source as $Q(t,x) = \delta(t) \cos(qx)$ and seek a solution for the substrate temperature in the form $T(t,z) \cos(qx)$. After applying this ansatz and performing a Fourier transform with respect to time, the thermal diffusion equation in the substrate takes the form

$$-i\omega \tilde{T} = -\alpha_s q^2 \tilde{T} + \alpha_s \frac{\partial^2 \tilde{T}}{\partial z^2} \quad (\text{S15})$$

with the following boundary condition at $z=0$:

$$-i\omega \sigma \tilde{T} = -q^2 \sigma \alpha_f \tilde{T} + \lambda_s \frac{\partial \tilde{T}}{\partial z} + 1, \quad z = 0 \quad (\text{S16})$$

where $\tilde{T}(\omega, t)$ is the Fourier transform of $T(z, t)$ with respect to time. From Eq. (S15) we get

$$\tilde{T} = \tilde{T}_0 \exp(-\beta z), \quad \beta = \left(q^2 - \frac{i\omega}{\alpha_s} \right)^{1/2} \quad (\text{S17})$$

where the branch of the square root should be chosen as to ensure the positive real part of β . Now from Eq.(S16) we get

$$T_0 = \frac{1}{-i\omega + \sigma \alpha_f q^2 + \beta \lambda_s} \quad (\text{S18})$$

The temperature of the film will be given by the inverse Fourier transform:

$$T(t, z = 0) = \frac{1}{2\pi} \int \frac{\exp(-i\omega t)}{-i\omega \sigma + \sigma \alpha_f q^2 + \lambda_s \left(q^2 - \frac{i\omega}{\alpha_s} \right)^{1/2}} d\omega \quad (\text{S19})$$

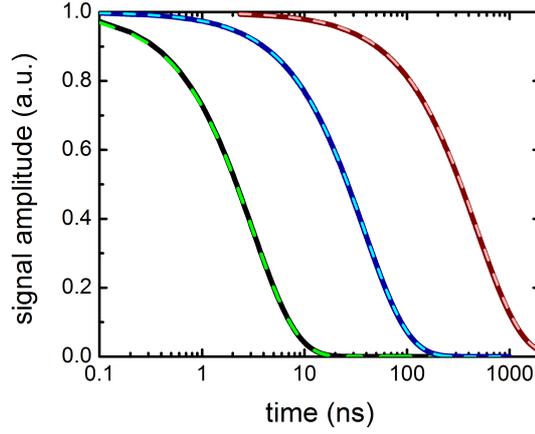


FIG. S5. Diffusive thermal decay of a silicon membrane in air (Eq. (S19), dashed lines) versus in vacuum (Eq. (1), heavy solid lines) for three representative grating periods: 2, 7, and 25 microns from left to right.

Figure S5 shows a comparison between numerical solutions to Eq. (S19) and exponential decay with the decay rates given by Eq.(1) for a silicon membrane in air. One can see that that the effect of the heat exchange with ambient air on the thermal grating relaxation is negligible. We found that one needs to increase the thermal conductivity of air in the model by a factor of 100 in order to get a noticeable effect.

Details of the effective thermal conductivity calculations

By changing the integration variable in Eq. (2), thermal conductivity can be represented as an integral over MFP,

$$k = \int_0^{\infty} k_{\Lambda} d\Lambda , \quad (\text{S20})$$

with the differential thermal conductivity $k_{\Lambda} = \frac{1}{3} C_{\omega} v \Lambda (d\Lambda/d\omega)^{-1}$ readily found from the thermal conductivity accumulation data with respect to MFP presented by Henry and Chen [S7]. The effective thermal conductivity in the transient grating geometry is found by multiplying the integrand by the correction factor,

$$k_{\text{eff}} = \int_0^{\infty} A(q\Lambda) k_{\Lambda} d\Lambda . \quad (\text{S21})$$

The above equation was used to calculate the solid curve in Fig. 3(b).

For a thin membrane, we need to account for the effect of boundary scattering. In Fuchs-Sondheimer theory, the effective MFP reduced due to diffuse scattering at the surfaces of the membrane is given by [S8]

$$\Lambda' = \Lambda \Phi \left(\frac{d}{\Lambda} \right) \quad (\text{S22})$$

$$\Phi(\chi) = 1 - \frac{3}{8\chi} + \frac{3}{2\chi} \int_1^{\infty} \left(\frac{1}{t^3} - \frac{1}{t^5} \right) e^{-\chi t} dt ,$$

where d is the membrane thickness. The correction factor Φ behaves similarly to the factor A in that it approaches unity in the limit of small MFP ($\Lambda \ll d$) and drops off when the MFP exceeds d .

Equation (S22) can be modified to allow for partially specular reflections [S7]. However, using the specular reflection probability as a free parameter introduces considerable arbitrariness in the data analysis; moreover, the commonly used model with a constant “specularity parameter” independent of both frequency and incidence angle [S8] is rather unphysical. Most studies of thermal conductivity in thin films use the “diffuse scattering” model, with satisfactory results [S9-S11].

To account for the combined effect of the heat transfer distance in the transient grating measurement and the boundary scattering in the membrane we take the MFP reduced by the boundary scattering from Eq. (S22) and plug it into Eq. (S21), which yields the following result:

$$k_{eff} = \int_0^\infty A(q\Lambda\Phi)k_\Lambda\Phi d\Lambda . \quad (S23)$$

This equation, with k_Λ from Henry and Chen [S7], was used to produce the theoretical curves for thin membranes in Fig. 3. Note that the wavevector dependence in Eq. (S23) is only present in the correction factor A resulting from non-diffusive transport. Equation (S23) is of course an approximation. For a rigorous analysis, one would need to solve the non-equilibrium thermal transport problem anew with appropriate boundary conditions at the surfaces of the membrane, which would be much harder than the analysis for an unbounded medium [S12].

- [S1] R.J. Collier, C.B. Burckhardt, and L.H. Lin. *Optical Holography* (Academic Press, New York 1971).
- [S2] L.M. Brekhovskikh. *Waves in Layered Media* (Academic Press, New York 1960).
- [S3] W. B. Gauster and D. H. Habing. *Phys. Rev. Lett.* **18**, 1058 (1967).
- [S4] C-M. Li, T. Sjodin, and H-L. Dai. *Phys. Rev. B* **56**, 15252 (1997).
- [S5] C. J. Glassbrenner and G. A. Slack. *Phys. Rev.* **134**, A1058 (1964).
- [S6] H.J. Eichler, F. Massmann, E. Biselli, M. Glotz, L. Konetzke, and X. Yang. *Phys. Rev. B* **36**, 3247 (1987).
- [S7] A. Henry and G. Chen. *J. Comp. Theor. Nanosci.* **5**, 141 (2008).
- [S8] E.H.Sondheimer. *Adv. Phys.* **1**, 1 (1952).
- [S9] Y.S. Ju and K.E. Goodson. *Appl. Phys. Lett.* **74**, 3005 (1999).
- [S10] W. Liu, K. Etessam-Yazdani, R. Hussin, and M. Asheghi. *IEEE Trans. Electron. Dev.* **53**, 1868 (2006).
- [S11] J.E. Turney, A.J.H. McGaughey, and C.H. Amon. *J. Appl. Phys.* **107**, 024317 (2010).
- [S12] A.A. Maznev, J. A. Johnson, and K.A. Nelson. *Phys. Rev. B* **84**, 1952061(2011).