

Electronic supplementary information

Structure relaxation and lattice mismatch

As the band structure of phosphorene is very sensitive to strain^[1], during the geometry optimization, we fixed the in-plane lattice parameters of the heterostructure to those of phosphorene ($a = 10.02 \text{ \AA}$, $b = 9.18 \text{ \AA}$). As a result of relaxation, there is no strain applied along the cross-plane direction. Finally, all the forces on each atom are relaxed within 0.01 eV/Å. After relaxation, the xy plane of gold (111) layers have strains of 1.5% and 4.3% along x - and y -axis, respectively. The graphene layer also has biaxial strains of 1.6% and 7.5% along two axes. The work functions of the gold (111) surface and graphene with the configuration in the heterostructure are calculated to be 5.1 eV and 4.6 eV, respectively, in good agreement with experiments. So we believe that the relaxed structure is reasonable when comparing with experiments.

The work function of gold is calculated by extracting the atomic coordinates of gold layers in the heterostructure. A vacuum of 10 Angstrom is added on top of gold layers. The Fermi energy E_F and the electrostatic potential E_{vac} in the middle of the vacuum are extracted from a self-consistent calculation. Then, the work function W is obtained as $W = E_F - E_{\text{vac}}$. The ionization potential of phosphorene is calculated in the same way but replacing the Fermi energy by the energy of valence band maximum E_{VBM} .

Wannier function calculations

Seven Wannier functions (WFs) with 5 atomic centered d orbitals and 2 σ orbitals centered on tetrahedron-interstitials (t orbital)^[2] are used to represent the gold atoms at the interface with graphene, while one t orbital is omitted for the inner gold atoms. Four atomic centered sp^3 orbitals are used for constructing the WFs of each phosphorus atom. For carbon atoms, there are two kinds of WFs with σ orbitals centered on mid-bond points and atomic centered p_z orbitals. After wannierization, the spatial spread of each WF is reduced to 15 bohr².

Electron transport.

To investigate the transport properties of the heterostructure, we use the real space Green's function method with the localized-orbitals basis Hamiltonian constructed by using maximally localized Wannier functions (MLWFs). The retarded Green's function of the heterostucture is:

$$\mathbf{G}^r = [(E + i\eta) \mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma}_L - \boldsymbol{\Sigma}_R]^{-1} \quad (\text{S1})$$

where \mathbf{I} the identity matrix, E the electron energy, $i\eta$ is a small imaginary part used to impose causality. $\boldsymbol{\Sigma}_\beta = \mathbf{H}_{C\beta} \mathbf{g}_\beta \mathbf{H}_{\beta C}$ ($\beta = L, R$) denotes the self-energy of the left (L) or right (R) leads, and \mathbf{g}_β is the surface Green's function of the leads, which can be calculated from the Hamiltonian matrix elements via an iterative procedure.^[3] The electron transmission function $\tau(E)$ can be calculated as:

$$\tau(E) = \text{Tr}(\mathbf{G}^r \boldsymbol{\Gamma}_L \mathbf{G}^a \boldsymbol{\Gamma}_R) \quad (\text{S2})$$

where $\mathbf{G}^a = (\mathbf{G}^r)^\dagger$ is the advanced Green's function and $\boldsymbol{\Gamma}_\beta = i(\boldsymbol{\Sigma}_\beta - \boldsymbol{\Sigma}_\beta^\dagger)$ ($\beta = L, R$). It should be noted that the above method is used for 1D transport. For the case of 3D transport, one uses translational invariance in the in-plane directions, i.e. the Bloch theorem, to describe the Hamiltonian as a sum over kpoints, in the 2D Brillouin zone perpendicular to the transport direction, of 1D Hamiltonians. For each k point, the k -dependent transmission $\tau(E, k)$ should be calculated and the total transmission would be $\tau(E) = \sum_k \tau(E, k) w_k$, where w_k is the weight of each transverse k point determined from the symmetry of the kpoint within the Brillouin zone. In the transmission calculation, we use a 30×30 k mesh to initialize the k -resolution and further interpolate^[4] to a 180×180 k grid to obtain a smooth transmission function.

By using the electron transmission function, one can obtain the coherent transport coefficients under the linear response approximation^[5, 6]:

$$G = -\frac{eJ}{\Delta\mu} \Big|_{\Delta T=0} = e^2 L_0$$

$$S = -\frac{\Delta\mu}{e\Delta T} \Big|_{J=0} = L_1 / (eTL_0)$$
(S3)

$$\kappa_{el} = -\frac{J_Q}{\Delta T} \Big|_{I=0} = (L_2 - L_1^2 / L_0) / T$$

$$L_n = 2/h \int dE \tau(E) \times (E - \mu)^n \times (-\partial f / \partial E)$$

where G , S , and κ_{el} denote the electron conductance, Seebeck coefficient, and electron thermal conductance, respectively. e is the charge of electron, and h is the Planck constant, μ is the chemical potential, T is the absolute temperature, and

$f(E, \mu, T) = 1 / (\exp[(E - \mu)/k_B T] + 1)$ is the Fermi-Dirac distribution function. Based on the electron transmission function, the 2-probe electrical current and electron thermal current can be calculated from the knowledge of the temperature and chemical potentials on both sides (L, R) as:

$$J = 2e/h \int dE \tau(E) (f_L - f_R) \quad (S4)$$

$$J_Q = 2/h \int dE \tau(E) (E - \mu) (f_L - f_R) \quad (S5)$$

Phonon transport.

First, we would like to clarify the labeling of the modes in graphene or phosphorene: The longitudinal and transverse modes of these layered structures, when isolated, have the polarization vector in the plane of G or P, and therefore do not contribute to thermal transport across these layers. Only the ZA modes with quadratic dispersion and polarization along the cross-plane direction, z, significantly contribute to the transport of heat. This labeling is the traditional way we refer to the modes of the isolated layer, and should not be confused with the way we label modes of the device. For instance a longitudinal mode of the device, has its momentum and polarization along z, while we are referring to the projection of this mode in the layers as ZA modes because they both have the *same polarization*.

The phonon cutoff frequency of gold is about 150 cm^{-1} , the intralayer optical phonon modes of phosphorene have a maximum frequency of about 470 cm^{-1} , while graphene has a maximum frequency of about 1600 cm^{-1} . As a result of this significant mismatch between the phonon spectra of gold and the layers, optical phonons of the latter, without considering phonon-phonon interaction, remain localized in the barrier region and only modes of frequency lower than 150 cm^{-1} can propagate through the device. Therefore only the acoustic modes of the layers, which have polarization along the transport direction z , namely the so-called ZA modes, would mostly contribute to the transport of heat. Thus, the contribution of intralayer optical phonons and the in-plane LA and TA modes are neglected in the thermal conductance calculation. This assumption has been validated in the case of 3 graphene layers between two gold leads, as displayed in Figure S3, where it was shown that the contribution of any mode other than ZA was more than 4 orders of magnitude smaller than that of the ZA mode.

Efficiency (COP) calculation.

The coefficient of performance (COP) is defined as the ratio of carried heat to the provided electrical energy:

$$\text{COP} = (J_Q - \kappa_{ph}\Delta T) / (JV) \quad (\text{S6})$$

To calculate the COP, we need to fix two of the four parameters ($J, J_Q, \Delta T, V$) and calculate the remaining two from equations (S4, S5) defining the currents in terms of applied electrical and thermal fields, and the knowledge of the transmission function. In practice, we fix the lead temperatures and chemical potentials (i.e. ΔT and $\Delta\mu$) and calculate the currents from (S4) and (S5). The reported COP was obtained by maximizing COP with respect to the applied bias at fixed ΔT . The applied bias V goes into the Fermi-Dirac distribution of the two electrodes: we set the chemical potential of the left electrode as $\mu_0 - e \times V/2$, and that of the right electrode as $\mu_0 + e \times V/2$. It turns out, as we also added in text, that the optimal bias is on

the order of few meVs, which implies there is hardly any shift in the energy levels under bias, and therefore one is in the linear regime where the transmission function is that of zero bias and zero temperature difference with a very good approximation.

- [1] A. S. Rodin, A. Carvalho, A. H. Castro Neto, Phys. Rev. Lett. **2014**, *112*, 176801.
- [2] I. Souza, N. Marzari, D. Vanderbilt, D. Phys. Rev. B **2002**, *65*, 035109.
- [3] M. P. L. Sancho, J. M. L. Sancho, J. Rubio, J. Phys. F: Met. Phys. **1985**, *15*, 851-858.
- [4] J. T. Falkenberg, M. Brandbyge, Beilstein J. Nanotechnol. **2015**, *6*, 1603–1608.
- [5] U. Sivan, Y. Imry, Phys. Rev. B **1986**, *33*, 551-558.
- [6] K. Esfarjani, M. Zebarjadi, Y. Kawazoe, Phys. Rev. B **2006**, *73*, 085406.

Φ (eV/ $\text{\AA}^2/\text{amu}$)	Au-Au	Au-G	G-G
zz	0.0383	0.0057	0.0168
xx or yy	0.00337	9.14e-5	0.000243

Table S1. Interlayer force constants between gold-gold, gold-graphene and graphene-graphene layers, obtained by moving one layer rigidly by 0.2 Ang and calculating the net restoring force on that layer as explained in the Equation (6) of the main text. As can be noted the zz components are much larger, justifying why in-plane modes have been neglected.

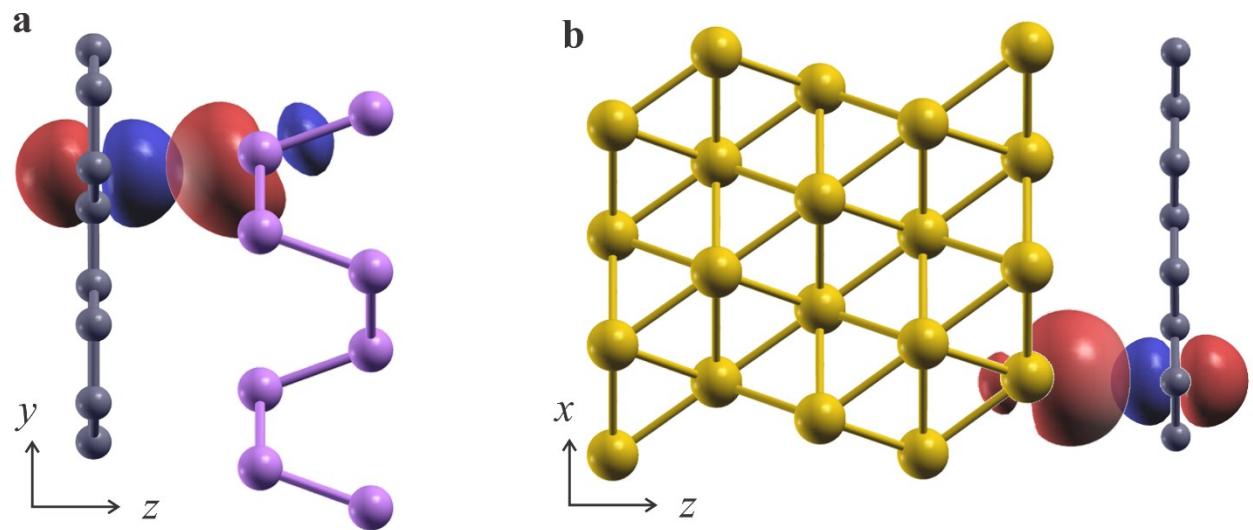


Figure S1. (a) p_z Wannier function (WF) of graphene overlaps with sp₃ WF of phosphorene.

(b) Surface WF of gold overlaps with p_z WF of graphene.

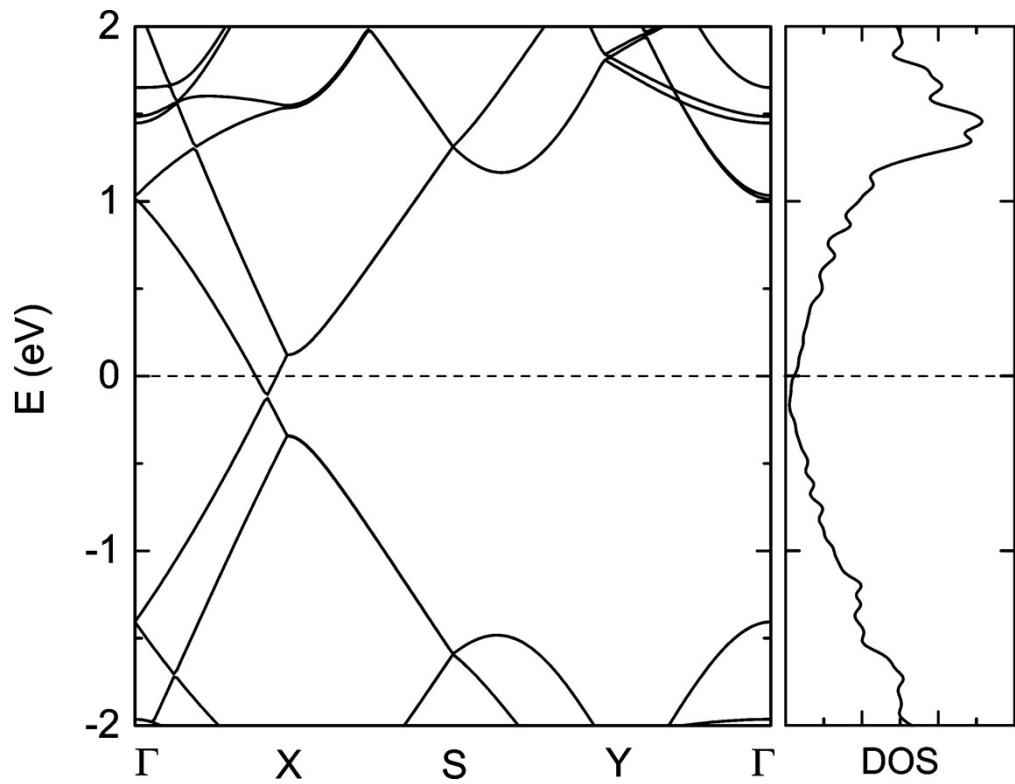


Figure S2. Local band structure and density of states (DOS) of graphene embedded in the heterostructure. The black dashed lines indicate the Fermi level.

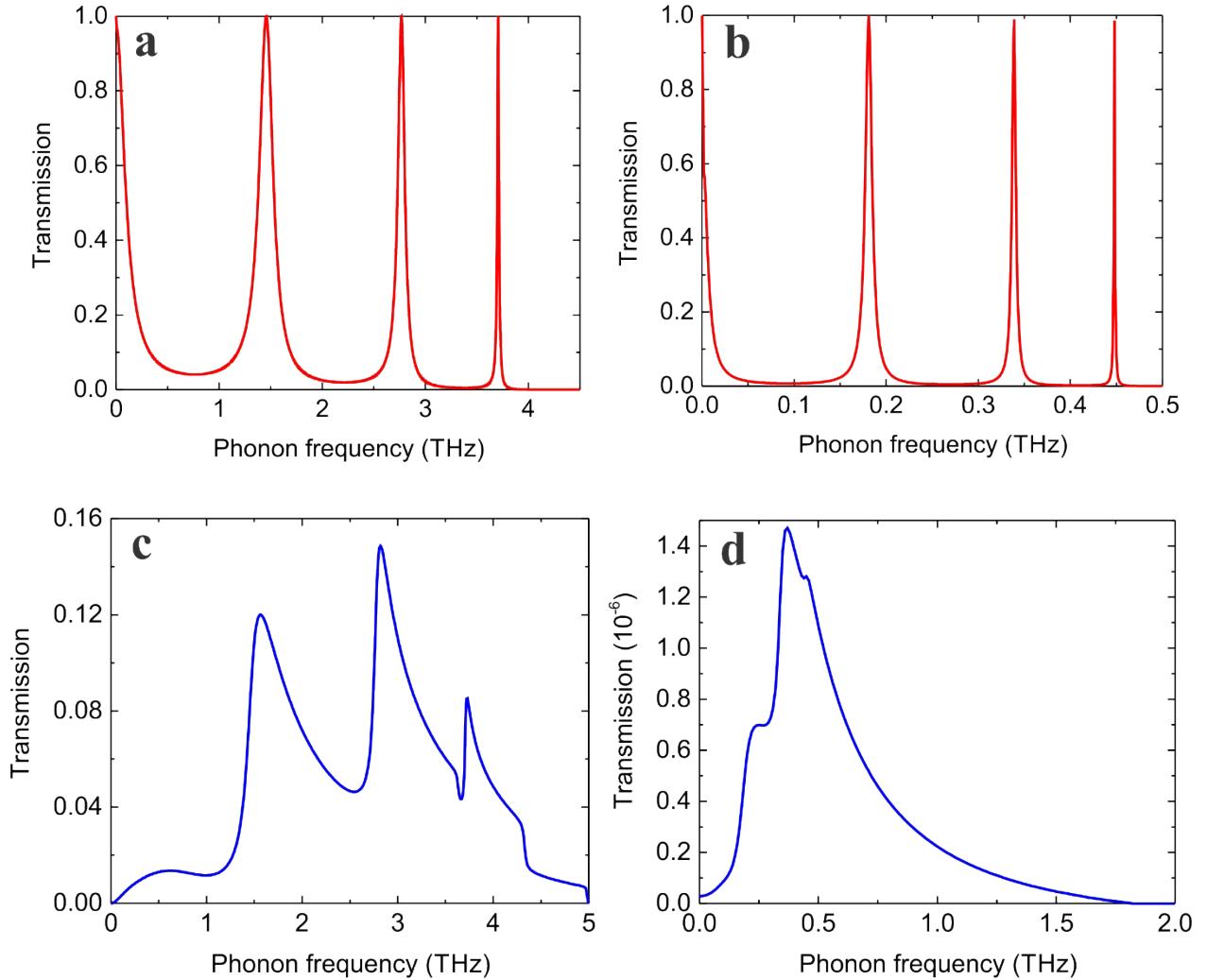


Figure S3. Phonon transmission function of Au-3G-Au system. (a) Modes polarized along the z direction (ZA modes) with $k = 0$ (normal incidence). (b) Modes with in-plane polarization and $k = 0$ (normal incidence). (c) Same as (a) but summed over all kpoints in the 2d Brillouin zone. (d) Same as (b) but summed over all kpoints in the 2d Brillouin zone. Note the transmission of the ZA mode is 5 orders of magnitude larger than the in-plane modes