Supplementary Figure 1 | Crystal structure and phonon dispersions of anatase TiO$_2$.

(a) Ball-and-stick model of the tetragonal unit cell of anatase TiO$_2$. Ti atoms are in gray and O atoms are in red. The lattice vectors and crystallographic directions referred to in the main text are indicated. (b) Calculated phonon dispersion relations of anatase TiO$_2$ along high-symmetry directions of the Brillouin zone. The $E_u$ and $A_{2u}$ phonons discussed in the main text are indicated. The $A_{2u}$ phonon is infrared active only along the $c$ axis, accordingly it is found to give a smaller contribution to the total Fröhlich coupling after integrating over all phonon momenta.
Supplementary Figure 2 | Non-adiabatic electron-phonon matrix elements in anatase TiO$_2$.

(a) Lindhard dielectric function associated with doped carriers in anatase TiO$_2$ at different values of $q$. For illustration we consider the highest doping level, $3.5 \times 10^{20}$ cm$^{-3}$.  (b) Non-adiabatic electron-phonon matrix elements $g^{NA}_{mn\nu}(k, q)$ corresponding to the LO E$_u$ phonon of anatase TiO$_2$, as a function of doping level. We set $n_k$ to the bottom of the conduction band. The electron-phonon matrix element becomes weaker at high doping. (c) Comparison between the density of states (DOS) near the conduction band bottom calculated from first principles (red) and modelled with a parabolic band (blue): $\text{DOS}(E) = V/(2\pi^2)(2m_b/\hbar^2)^{3/2}\sqrt{E}$, where $V$ is the unit cell volume of anatase. The ab initio DOS was computed using Wannier interpolation to obtain the electronic energies on a randomly generated dense $k$ grid and with a Gaussian smearing of 5 meV. The gray vertical line indicates the energy corresponding to the Fermi level at the highest doping. The parabolic DOS starts deviating significantly from the calculated one for energies above 0.2 eV.
Supplementary Figure 3 | Spectral function in the transition regime. (a) Spectral function of anatase TiO$_2$ calculated for the doping level $1 \times 10^{20}$ cm$^{-3}$. Gaussian masks of widths 25 meV and 0.015 Å$^{-1}$ were applied as in Fig. 1 of the main text. (b) Band structure extracted from (a) (blue lines) together with the bare band (red line). For this doping one satellite is still visible but the mass renormalization parameter is decreasing ($\lambda = 0.34$ as reported in the main text).
Supplementary Figure 4 | Impact of the electron lifetime on the spectral properties. Spectral function of anatase TiO$_2$ calculated for the doping level $3.5 \times 10^{20}$ cm$^{-3}$ using $\hbar/\tau = 25$ (a), 55 (b) and 75 meV (c) to compute the electronic screening (see Methods). The spectral features are virtually unchanged. We recomputed the mass renormalization parameter $\lambda$ as in the main text which gives $\lambda = 0.19$, 0.20 and 0.22 for increasing broadening. The same analysis for the doping concentration $5 \times 10^{18}$ cm$^{-3}$ yields $\lambda = 0.77$, 0.73 and 0.72 for the same broadenings. These values fall within 10% of the results presented in the main text.
Supplementary Note 1 | Cumulant expansion

The spectral function is related to the imaginary part of the one-electron retarded Green’s function by:

\[ A(k, \omega) = -\frac{1}{\pi} \sum_n \text{Im} G_{nk}(\omega). \]  

(1)

In the cumulant expansion the Green’s function is obtained in the time domain, starting from the interaction picture:

\[ G_{nk}(t) = i \exp \left[ -i\varepsilon_{nk} t/\hbar + C_{nk}(t) \right], \]  

(2)

where \( C_{nk}(t) \) is the cumulant function. By taking the Fourier transform of this expression to the frequency domain and inserting it in Supplementary Eq. (1) we obtain Eq. (1) of the main text. In order to obtain an expression for the cumulant which is amenable to computation, it is customary to expand the exponential in powers of the cumulant:

\[ G_{nk} = G_{nk}^0 \left[ 1 + C_{nk} + C_{nk}^2/2 + \cdots \right], \]  

where \( G_{nk}^0 = i \exp (-i\varepsilon_{nk} t/\hbar) \) is the Green’s function in absence of electron-phonon interactions. Alternatively, the Green’s function can be obtained from the Dyson equation.

\[ G_{nk} = G_{nk}^0 + G_{nk}^0 \Sigma_{nk} G_{nk}^0 + G_{nk}^0 \Sigma_{nk} \Sigma_{nk} G_{nk}^0 + \cdots. \]  

By comparing these expansions term-by-term one finds an explicit expression for the cumulant function, which is given in the Methods and is reproduced here for completeness:

\[ C_{nk}(t) = \frac{1}{\pi \hbar} \int_{-\infty}^{\infty} d\omega \frac{\text{Im} \Sigma_{nk}(\varepsilon_{nk}/\hbar - \omega)}{(\omega - i\eta)^2} e^{(i\omega + \eta)t} \theta(\varepsilon_F - \varepsilon_{nk} + \hbar \omega). \]  

(3)

In practical calculations the exact self-energy \( \Sigma_{nk} \) is replaced by the best available approximation, which is the Migdal expression given in Eq. (2) of the main text. A rigorous derivation of the cumulant expansion and a discussion of its advantages and limitations can be found in Supplementary Refs. 4, 5. In particular, in Supplementary Ref. 4 it is shown that the choice of seeding the self-energy calculated with the first non-crossing diagram also guarantees that no overcounting of correlated higher-order contributions is introduced in the theory. The inclusion of crossing diagrams in the evaluation of the Green’s function results from the time orderings of the \( t \) variables in the cumulant expansion.

The spectral function given in Eq. (1) of the main text yields multiple bosonic satellites, one for each term in the Taylor expansion of \( \exp[C_{nk}(t)] \). A convenient expression for the case of a
single satellite was derived in Supplementary Ref. 3. In this work we extended the method of Supplementary Ref. 3 to the case of multiple satellites. Following Supplementary Ref. 3, we write the spectral function as:

\[ A(k, \omega) = \sum_n \left[ A_{nk}^{QP}(\omega) + A_{nk}^{S1}(\omega) + A_{nk}^{QP}(\omega) * A_{nk}^{S2}(\omega) + \cdots \right], \]  

(4)

where * indicates the convolution. In the last expression the quasiparticle contribution \( A_{nk}^{QP}(\omega) \) is defined as:

\[ A_{nk}^{QP}(\omega) = \frac{2}{\pi} \frac{|\text{Im} \Sigma_{nk}(\varepsilon_{nk})|}{\left[ h\omega - \varepsilon_{nk} - \text{Re} \Sigma_{nk}(\varepsilon_{nk}) \right]^2} \]  

(5)

and the satellite contributions associated with one-phonon and two-phonon processes are:

\[ A_{nk}^{S1}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} C_{nk}(t), \quad A_{nk}^{S2}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \frac{1}{2} C_{nk}(t) C_{nk}(t). \]  

(6)

The function \( A^{S1} \) can be written in terms of the electron-phonon self-energy by combining Supplementary Eqs. (6) and (3) and carrying out the Fourier transform. This step was performed in Supplementary Ref. 3:

\[ A_{nk}^{S1}(\omega) = \beta_{nk}(\omega) - \beta_{nk}(\varepsilon_{nk}) - (\omega - \varepsilon_{nk}/\hbar) \frac{\partial \beta_{nk}}{\partial \omega} \bigg|_{\varepsilon_{nk}/\hbar}, \]  

(7)

with

\[ \beta_{nk}(\omega) = \frac{1}{\pi} \text{Im} \Sigma_{nk}(\varepsilon_{nk}/\hbar - \omega) \theta(\varepsilon_F/\hbar - \omega). \]  

(8)

In order to obtain the contribution of the second satellite, we note that \( A_{nk}^{S1}(\omega) \) is the Fourier transform of \( C_{nk}(t) \), therefore the expression for \( A_{nk}^{S2}(\omega) \) in Supplementary Eq. (6) can be rewritten using the convolution theorem:

\[ A_{nk}^{S2} = \frac{1}{2} A_{nk}^{S1} * A_{nk}^{S1}. \]  

(9)

By combining Supplementary Eqs. (9) and (4) we obtain the expression used in our calculations:

\[ A(k, \omega) = \sum_n \left[ A_{nk}^{QP}(\omega) + A_{nk}^{S1}(\omega) + A_{nk}^{QP}(\omega) * A_{nk}^{S1}(\omega) + \frac{1}{2} A_{nk}^{S1}(\omega) * A_{nk}^{S1}(\omega) * A_{nk}^{QP}(\omega) + \cdots \right]. \]  

(10)

From Supplementary Eq. (10) we can extrapolate the general expression for the case of many satellites:

\[ A(k, \omega) = \sum_n \left[ 1 + A_{nk}^{S1}(\omega) * + \frac{1}{2} A_{nk}^{S1}(\omega) * A_{nk}^{S1}(\omega) * + \cdots \right] A_{nk}^{QP}(\omega) = \sum_n e^{A_{nk}^{S1}(\omega)} A_{nk}^{QP}(\omega). \]  

(11)
When considering a drastically simplified model system consisting of dispersionless electrons and an Einstein phonon spectrum, this last expression reduces to the well-known Lang-Firsov series of polaron satellites\textsuperscript{1,6}. In fact, by setting $A^{QP} = Z \delta(\omega)$ and $A^{S1} = \lambda \delta(\omega - \omega_{\text{ph}})$, Supplementary Eq. (11) gives $A(\omega) = Z \sum_{m=0}^{\infty} (\lambda^m / m!) \delta(\omega - m \omega_{\text{ph}})$. By further requiring the normalization of the spectral function, that is $\int A(\omega) d\omega = 1$, we obtain $Z = e^{-\lambda}$ and the Lang-Firsov expression is recovered. In our \textit{ab initio} calculations we truncate Supplementary Eq. (11) to the second order. In practice we first evaluate the quasiparticle and satellite contributions, and then we perform two successive numerical convolutions in order to obtain the two satellites which are seen in the experiments.

**Supplementary Note 2 | Mass enhancement and Fröhlich coupling constant**

The mass enhancement parameters obtained from the quasiparticle bands along the $\Gamma \Sigma$, $\Gamma X$, and $\Gamma Z$ directions are $\lambda = 0.73$, 0.73, and 0.74 at the lowest doping; $\lambda = 0.70$ in each direction at the intermediate doping; and $\lambda = 0.20$, 0.20, and 0.18 at the highest doping. The mass enhancement can also be calculated directly from the energy derivative of the real part of the electron self-energy at the Fermi level, $\lambda_k = -\partial \text{Re} \Sigma_k(\omega)/\partial \omega|_{\omega=\epsilon_F}$. We checked that the values thus obtained fall within less than 10% of the ones listed above, the average over the Fermi surface being $\lambda = 0.68$, 0.65 and 0.19 for the three values of doping considered. This approach also validates our calculations of the contribution to $\lambda$ arising from different phonon modes presented in the analysis of Fig. 2a in the main text.

From the mass enhancement parameter it is also possible to obtain the quasiparticle strength $Z$ as $Z = 1/(1 + \lambda)^{\text{7}}$. Using the values of $\lambda$ reported in Fig. 1g-i we obtain $Z = 0.58$, 0.59, and 0.83 with increasing doping. Our calculated quasiparticle strength at intermediate doping overestimates the experimentally-determined value, $Z = 0.36^{\text{8}}$. We attribute this difference to extrinsic losses not accounted for in our calculations; these losses are known to transfer spectral weight to the satellites\textsuperscript{3,9,10}.

In polar materials it is customary to describe the coupling to a dispersionless LO phonon via the
dimensionless Fröhlich coupling constant $\alpha$:

$$\alpha = \frac{e^2}{\hbar} \left( \frac{m_b}{2\hbar \omega_{LO}} \right)^{1/2} \left( \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0} \right),$$

where $m_b$ is the band mass, $\hbar \omega_{LO}$ the energy of the LO phonon, $\epsilon_0$ and $\epsilon_\infty$ the static and high-frequency permittivities, respectively. We evaluate Supplementary Eq. (12) using the experimental values of $\omega_{LO}$, $\epsilon_0$, and $\epsilon_\infty$ from Supplementary Ref. 11. For the effective masses we use our DFT calculations since no accurate experimental values are available: $m_\perp = 0.40 m_e$ and $m_\parallel = 4.03 m_e$. By using these parameters in Supplementary Eq. (12) we find the experimental Fröhlich constants $\alpha_\perp = 1.0$ and $\alpha_\parallel = 3.4$, and their isotropic average $\alpha = 1.8$. The Fröhlich constant $\alpha$ can also be obtained from \textit{ab initio} calculations of the electron-phonon matrix elements, following Supplementary Ref. 12. In this case we calculate $\alpha_\perp = 0.9$, $\alpha_\parallel = 3.0$, and the isotropic average $\alpha = 1.6$. These values are in excellent agreement with experiment, therefore our description of Fröhlich coupling in anatase TiO$_2$ is expected to be highly accurate. We note that the anisotropy of the Fröhlich coupling constant $\alpha_\perp$ and $\alpha_\parallel$ does not stem from anisotropic electron-phonon interactions, but rather from the strong anisotropy in the band masses, which enter $\alpha$ as seen in Supplementary Eq. (12).

For weak and intermediate couplings, the polaron effective mass is commonly estimated using $m^* = m_b (1 - \alpha/6)^{-1}$. While this procedure is adequate for isotropic crystals, it cannot be used in the present case of TiO$_2$. In fact, if we use the isotropic average of the coupling constant, then we underestimate $m^*$ with respect to experiment. On the other hand, if we consider $\alpha_\perp$ and $\alpha_\parallel$ separately, then we obtain a strongly anisotropic mass enhancement, which is not consistent with our many-body calculations. These observations indicate that, in the case of anisotropic crystals, the Fröhlich constant $\alpha$ should be used with caution.

**Supplementary Note 3 | Dielectric screening**

In this work we calculate the additional screening arising from the charge carriers using the random phase approximation (RPA) for the homogeneous electron gas, that is the Lindhard screening. This choice is motivated by the fact that we must evaluate the screening for millions of electron-phonon matrix elements, and explicit \textit{ab initio} calculations of the RPA screening for such a dense Brillouin
zone grid are not currently feasible. The use of the Lindhard model is justified by the fact that the system under consideration lies in the high-density electron-gas limit. To confirm this point we evaluate the Wigner-Seitz radius \( r_s \) given by
\[
r_s = \left( \frac{4\pi n a_0^*}{3} \right)^{-1/3},
\]
where \( n \) is the doping density and \( a_0^* = \frac{\hbar^2}{e^2\epsilon_0} \). Using the band effective mass \( m_e \) and the static dielectric constant \( \epsilon_0 \) of anatase we obtain \( r_s = 1.6 \) for the lowest doping level \( 5 \times 10^{18} \text{ cm}^{-3} \). This value is comparable or even smaller than what found in simple metals, therefore the use of the electron gas model to describe doped carriers is justified. Moreover, the electronic bands are to a good approximation parabolic in the energy range considered in this work; this is seen by comparing the \textit{ab initio} density of states with the parabolic band model, Supplementary Fig. 2c. Therefore the use of a Lindhard function is justified, and we expect this choice to be very accurate in the present case.

**Supplementary Note 4 | Polaron wavefunction**

In order to calculate the wavefunction of a polaron, we follow the approach of Supplementary Ref. 15 and express the many-body electron-phonon state using Rayleigh-Schrödinger perturbation theory. We consider a system at zero temperature and with a single electron added to the bottom of the conduction band. The resulting expression is:

\[
\tilde{\psi}_{nk}(r; \{\tau_\kappa\}) = \psi_{nk}(r)|0_p\rangle + \frac{1}{\sqrt{N_q}} \sum_{mnq} g_{mnq}(k, q) \psi_{mk+q}(r) \hat{a}_{-\nu}^+ |0_p\rangle ,
\]

where \{\( \tau_\kappa \)\} are the nuclear coordinates, \( \hat{a}_{-\nu}^+ \) is a phonon creation operator, and \( |0_p\rangle \) is the ground state with no phonons. The Brillouin zone is discretized using a uniform grid with \( N_q \) phonon wavevectors. Since the atomic displacements are smaller than characteristic interatomic distances, we can simplify the above expression by replacing \( \hat{a}_{-\nu}^+ \) (which is a function of the normal mode coordinates) by its average over a given electron-phonon state\(^{15}\). In order to determine a lower bound to the polaron radius, we evaluate this expectation value by considering an electronic wavefunction localized at the center of the reference frame, \( r = 0 \). The result is:

\[
\langle \hat{a}_{-\nu}^+ \rangle = \frac{1}{\sqrt{N_q}} \sum_{m'} \frac{g_{m'nq}^*(k, q)}{\epsilon_{nk} - \epsilon_{mk+q} - \omega_{\nu}}.
\]
By replacing $\langle \hat{a}^\dagger_{-q\nu} \rangle$ inside Supplementary Eq. (13) we have:

$$\tilde{\psi}_{nk}(r; \{\tau_n\}) = \frac{e^{ik \cdot r}}{\sqrt{\Omega}} \left[ u_{nk}(r) + \frac{1}{N_q} \sum_{m\nu} g_{mn\nu}(k, q) u_{mk+q}(r) e^{i\mathbf{q} \cdot r} \sum_{m'} \frac{g^*_{m'n\nu}(k, q)}{\varepsilon_{nk} - \varepsilon_{m'k+q} - \omega_{q\nu}} \right] |0_p\rangle.$$  

(15)

The expression for $\tilde{\psi}_{nk}$ given in the Methods was obtained from Supplementary Eq. (15) by considering the normalization of the wavefunction and by retaining the long-wavelength part of the Fröhlich vertex, so that $g_{mn\nu}(k, q) \to g_{mn\nu}(k, q) \delta_{mn}$. We already demonstrated that this approximation is very accurate for anatase TiO$_2^{12}$. Since the main contribution to Supplementary Eq. (13) arises from long-wavelength optical phonons, we also replace $u_{mk+q}(r)$ by $u_{mk}(r)$, in the spirit of the $k \cdot p$ approximation.

Our choice of calculating the polaron wavefunction within Rayleigh-Schrödinger perturbation theory is justified since this theory is valid in the range $\alpha \leq 5$ (see Supplementary Ref. 15, p. 516), and the largest Fröhlich coupling constant in anatase TiO$_2$ is $\alpha_{\parallel} = 3.0$ (see Supplementary Note 2).

We note that the present many-body approach for calculating the polaron wavefunction differs significantly from DFT calculations of excess electrons in oxides using large supercells$^{16-19}$. In fact in the present case we employ a quantum description of both electrons and nuclei, and a dynamical (non-adiabatic) description of their interactions. In contrast, supercell DFT calculations describe the nuclei as classical particles and decouple electronic and nuclear degrees of freedom using the Born-Oppenheimer adiabatic approximation. These approximations are not justified in the case of TiO$_2$, and have led to results which are very sensitive to the DFT exchange and correlation functional and the size of the supercell. Our present perturbative approach does not suffer from these shortcomings.
Supplementary References


