

Hidden self-energies as origin of cuprate superconductivity revealed by machine learning

Youhei Yamaji,^{1,2,*} Teppei Yoshida,³ Atsushi Fujimori,⁴ and Masatoshi Imada¹

¹*Department of Applied Physics, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*

²*JST, PRESTO, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*

³*Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501, Japan*

⁴*Department of Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan*

(Dated: March 20, 2019)

Experimental data are the source of understanding matter. However, measurable quantities are limited and theoretically important quantities are often hidden. Nonetheless, recent progress of machine-learning techniques opens possibilities of exposing them only from available experimental data. In this article, the Boltzmann-machine method is applied to the angle-resolved photoemission spectroscopy spectra of cuprate superconductors. We find prominent peak structures both in normal and anomalous self-energies, but they cancel in the total self-energy making the structure apparently invisible, while the peaks make dominant contributions to superconducting gap, hence providing a decisive testimony for the origin of superconductivity. The relation between superfluid density and critical temperature supports involvement of universal carrier relaxation time associated with dissipative strange metals. The present achievement opens avenues for innovative machine-learning spectroscopy method.

I. INTRODUCTION

Momentum k and energy ω dependent electron single-particle spectral function $A(k, \omega)$ can be measured with recent revolutionarily refined resolution of angle resolved photoemission spectroscopy (ARPES)¹. From $A(k, \omega)$, the interaction effects crucial for unconventional superconductors can be identified in the self-energy^{2,3}. Scanning tunnel microscope (STM) and its spectra (STS) including the quasiparticle interference method⁴ also give us insights into the self-energy^{5,6}.

In superconductors, the self-energy consists of the normal and anomalous (superconducting) contributions, Σ^{nor} and Σ^{ano} , respectively. ARPES and STS provide us with only the total self-energy Σ^{tot} in a specific combination of these two² (see below for details). However, to understand the superconducting mechanism, it is crucially important to extract these two separately, because they represent theoretically different part of interaction effects: Σ^{ano} is proportional to the superconducting gap function, at the heart of superconducting properties, while normal-electron correlation effects, such as renormalized mass and life time, are encoded in Σ^{nor} . Despite its importance, Σ^{ano} can be straightforwardly extracted separately only when Σ^{nor} is non-singular as in the BCS superconductivity of weakly correlated systems^{5,6}.

In this paper, we develop a scheme of machine-learning technique to extract physical quantities hidden in experimental data. To demonstrate the power of our method, we apply it to a long-standing open issue of the electronic structure in the cuprate high-temperature superconductors under strong correlation effects manifested by the formation of the pseudogap in the normal state. Specifically, the Boltzmann machine^{7,8} is examined to extract Σ^{nor} and Σ^{ano} separately from available ARPES spectra

even when the normal self-energy is subject to prominent or singular correlation effects.

II. EXPERIMENTAL DATA AND THEORETICAL SETUP

We utilize a high resolution ARPES data taken for two cuprate compounds, $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (Bi2212) for optimum doping concentration with critical temperature $T_c \sim 90\text{K}$ ⁹ and $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$ (Bi2201) as an under-doped example ($T_c \sim 29\text{K}$)¹⁰. We analyse Bi2212 data at temperature $T = 12\text{ K}$ and Bi2201 at $T = 11\text{K}$, which are both well below T_c . The machine learning enables us to obtain Σ^{nor} and Σ^{ano} separately, and reveals prominent peak structures in both of them, which are, however, apparently hidden in the original ARPES data, because of the cancellation of these two contributions. We elucidate its profound consequences for the superconducting mechanism.

More concretely, the momentum k and energy ω dependent spectral function $A(k, \omega) = -(1/\pi)\text{Im}G^{\text{nor}}(k, \omega)$ is a functional of Σ^{nor} and Σ^{ano} , where the normal component of the Green's function is given as

$$G^{\text{nor}}(k, \omega)^{-1} = \omega - \epsilon_k - \Sigma^{\text{tot}}(k, \omega). \quad (1)$$

with bare-electron dispersion ϵ_k . Here, $\Sigma^{\text{tot}}(k, \omega) = \Sigma^{\text{nor}}(k, \omega) + W(k, \omega)$ is given by a specific combination of Σ^{nor} and Σ^{ano} , where the anomalous contribution $W(k, \omega)$ is given by

$$W(k, \omega) = \Sigma^{\text{ano}}(k, \omega)^2 / [\omega + \epsilon_k + \Sigma^{\text{nor}}(k, -\omega)^*], \quad (2)$$

in the retarded representation¹¹. The gap function

$$\Delta(k, \omega) = Q(k, \omega)\Sigma^{\text{ano}}(k, \omega), \quad (3)$$

which is a measure of superconducting order, is proportional to $\Sigma^{\text{ano}}(k, \omega)$ with the coefficient $Q(k, \omega)$ called the renormalization factor (see Appendix A).

Although tremendous efforts have been devoted since the discovery of the cuprate superconductors with many fruitful clarifications, various puzzling issues remain open. The normal-state $A(k, \omega)$ is largely unusual including the pseudogap. Nevertheless, the superconducting phase does not look unusual except for the d -wave-type nodal gap itself and somewhat inconspicuous “peak-dip-hump” structure (see red square symbols in Fig. 1a.^{9,10}): Outside the sharp quasiparticle peak (at ~ -40 meV in Fig. 1a.) expected at the superconducting gap edge, $A(k, \omega)$ (energy distribution curve (EDC)) particularly at the antinodal point $k = k_{\text{AN}}$ is characterized by a deeper-energy weak dip followed by a broad hump^{1,2}. In contrast, the underdoped sample does not show the gap-edge peak (Fig. 1b.), though comparable gaps ~ 30 meV open as a first look. They are in contrast with the strong-coupling BCS superconductors, where the phonons generate prominent peaks (or saw-tooth-like) structures outside the gap in $A(k, \omega)$ (or density of states after angle integration), which is identified as the major origin of the superconducting gap and is the decisive testimony of the electron-phonon mechanism^{5,6,12}.

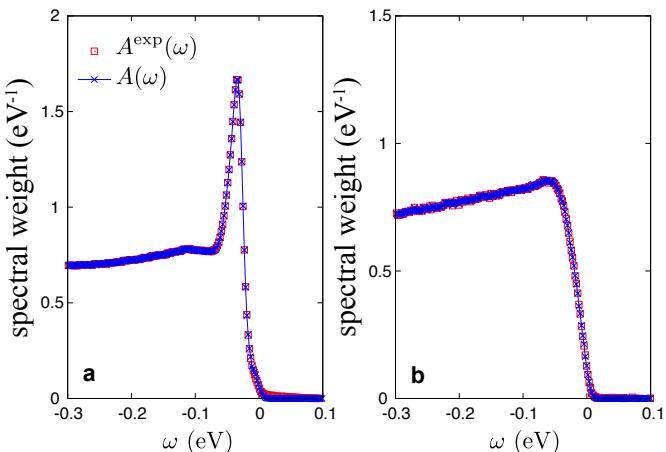


FIG. 1. Comparison of experimental $A(k_{\text{AN}}, \omega)$ and Boltzmann-machine fitting. Spectral weight (EDC curve) $A(k_{\text{AN}}, \omega)$ of **a.** optimally hole-doped $\text{Bi}_2\text{Sr}_2\text{Ca}\text{Cu}_2\text{O}_{8+\delta}$ (Bi2212) at 12 K (left) (Ref. 10) and **b.** underdoped $\text{Bi}_2\text{Sr}_2\text{Cu}\text{O}_{6+\delta}$ (Bi2201) at 11 K (right) (Ref. 10) at the antinodal point $k = k_{\text{AF}}$ (more precisely the closest point to the antinodal point, at which the momentum distribution of the quasiparticle dispersion curve is peaked). Red squares are experimental data. Blue crosses are reconstructed from the self-energies, which are deduced from Boltzmann machines.

III. HIDDEN SELF-ENERGY STRUCTURE REVEALED BY BOLTZMANN MACHINE

By using the Boltzmann-machine learning, a dramatic consequence is revealed for Σ^{nor} and Σ^{ano} by reconstructing them from the mild structure of $A(k, \omega)$ given by ARPES. The present reconstruction is a non-linear underdetermined problem as in many of machine learning problems. To obtain a reliable solution, we utilize physically sound constraints such as the rigorous causality encoded as the Kramers-Kronig relation. Sparse and localized nature of $\text{Im}\Sigma^{\text{ano}}$ is resulted *a posteriori* as the optimized solution under physical constraint as detailed in Appendix A. To represent the self-energies and incorporate the physical constraints, the Boltzmann machine as universal function approximators developed in machine learning is employed. See also Fig. 5 for the flow chart.

The obtained $A(k_{\text{AN}}, \omega)$ (cross points in Fig. 1) perfectly reproduces the distinct behaviors of both of the optimally doped and underdoped samples. These EDC curves are constructed from $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ in Figs. 2 **a.** and **b..** Remarkably, prominent peaks are found in $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ at $\omega = \omega_{\text{PEAK}}^{\text{OP}} \sim \pm 50$ meV for Bi2212 and at $\omega = \omega_{\text{PEAK}}^{\text{UD}} \sim \pm 45$ meV for Bi2201, accompanied by weaker peaks at ± 180 meV and ± 160 meV, respectively. We will show later that the discovered peaks are the main source of superconductivity. Although the peak of $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ had been searched for long time in analogy to the strong coupling BCS superconductors^{2,13}, its clear signature was missing in the cuprates. The machine learning now has succeeded in its identification. Surprisingly, $\text{Im}\Sigma^{\text{nor}}(\omega)$ also has distinct (positive or negative) peaks at the same energy as $\text{Im}\Sigma^{\text{ano}}(\omega)$ and as we later clarify their contributions to the spectral function cancel. The robustness of our finding against noise and the experimentally uncertain high-energy part is clarified in Appendix B. We note that in the case of underdoped Bi2201 sample below T_c , there exists subtlety in the machine learning solution. Although the superconducting solution presented here is the optimized solution with the smallest mean-square error in the fitting of $A(k, \omega)$, an insulating solution is also found with larger error. This may be related to severe competition of insulating and superconducting behaviors in the real sample. We show the properties of superconducting solution because it gives the best optimized solution and the sample is indeed superconducting.

Despite the peaks in $\text{Im}\Sigma^{\text{nor}}$, and $\text{Im}\Sigma^{\text{ano}}$, prominent peaks are missing in $\text{Im}\Sigma^{\text{tot}}(\omega)$ as shown in Figs. 2**c.** and **d..** We discuss below why the peaks in $\text{Im}\Sigma^{\text{ano}}$ at $\omega \neq 0$ necessarily show up and their contribution cancels with $\text{Im}\Sigma^{\text{nor}}$, when we impose physical constraints such as the Kramers-Kronig transformation. Instead of the peak in $\text{Im}\Sigma^{\text{nor}}(\omega)$ and $\text{Im}\Sigma^{\text{ano}}(\omega)$, a negative prominent peak¹⁴ generating the superconducting gap is found centered at $\omega \sim 0$ in $\text{Im}\Sigma^{\text{tot}}$, which arises from the zero of the denominator in Eq.(2), commonly to the conventional BCS superconductors.

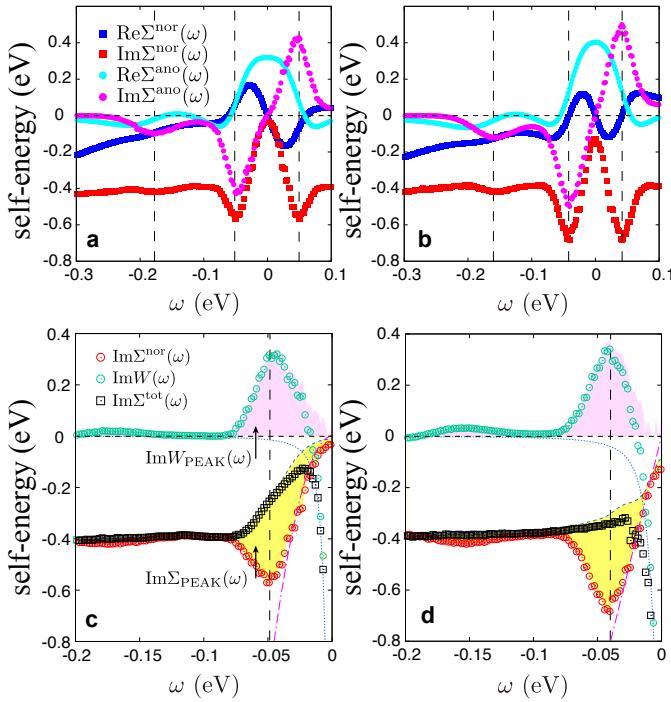


FIG. 2. Normal and anomalous self-energies derived from machine learning and their decomposition. Normal self-energy $\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and anomalous self-energy $\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ deduced from $A(k_{\text{AN}}, \omega)$ in Fig. 1 by the machine learning for Bi2212 (**a**) and Bi2201 (**b**). The vertical dashed lines indicate the peak positions ω_{PEAK} in the imaginary part. $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$, $\text{Im}W(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{tot}}(k_{\text{AN}}, \omega)$ are plotted for Bi2212 (**c**) and Bi2201 (**d**). The peaks of $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and $\text{Im}W(k_{\text{AN}}, \omega)$ are completely canceled in their sum $\text{Im}\Sigma^{\text{tot}}(k_{\text{AN}}, \omega)$. $\text{Im}\Sigma^{\text{tot}}$ is decomposed into simple BCS-type superconducting contribution $L_{\text{BCS}}(k_{\text{AN}}, \omega)$ (a Lorentzian around $\omega = 0$ (dotted blue curve)) and the rest $\text{Im}\Sigma_N(k_{\text{AN}}, \omega) \equiv \text{Im}\Sigma^{\text{tot}} - L_{\text{BCS}}(k_{\text{AN}}, \omega)$. The latter is fitted by a superposition of many Gaussian distributions (blue dashed curve). Then the unusual structures are identified as $\text{Im}\Sigma_{\text{PEAK}}(k_{\text{AN}}, \omega) \equiv \text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega) - \text{Im}\Sigma_N(k_{\text{AN}}, \omega)$ (yellow shaded area) and $\text{Im}W_{\text{PEAK}}(k_{\text{AN}}, \omega) \equiv \text{Im}W(k_{\text{AN}}, \omega) - L_{\text{BCS}}(k_{\text{AN}}, \omega)$ (pink shaded area). The yellow and pink areas cancel in their sum both in **c**. and **d**. The magenta dash-dotted curves show a quadratic (linear) fitting of $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ of Bi2212 (Bi2201) for $|\omega| < 35$ meV.

To understand the significance of the peaks at ω_{PEAK} in $\text{Im}\Sigma^{\text{ano}}$, we show the contribution of the peaks to $\text{Re}\Sigma^{\text{ano}}(k, \omega = 0)$ estimated from the normalized partial Kramers-Kronig relation¹⁵ (Cauchy relation) (see Appendix A),

$$I_{\Sigma}(\Omega) = \frac{2 \int_0^{\Omega} d\omega |\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)| / \omega}{\pi \text{Re}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega = 0)}. \quad (4)$$

$\text{Re}\Sigma^{\text{ano}}(k, \omega = 0)$ is a measure of the superconducting amplitude, because the gap $\Delta(k, \omega = 0)$ is proportional to $\Sigma^{\text{ano}}(k, \omega = 0)$ (Eq.(3)). Since $I_{\Sigma}(\Omega = \infty) = 1$,

the contribution of the peak in $\text{Im}\Sigma^{\text{ano}}$ to the superconductivity can be estimated from the increment in $I_{\Sigma}(\Omega)$. Figure 3a. shows that the inner energy peak at $\omega_{\text{PEAK}} = -50$ meV (-45 meV) for Bi2212 (Bi2201) both contribute to more than 90% of $\text{Re}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega = 0)$ (note that $\text{Im}\Sigma^{\text{ano}}$ is an odd function of ω). Namely, these peaks are the main origin of the superconductivity.

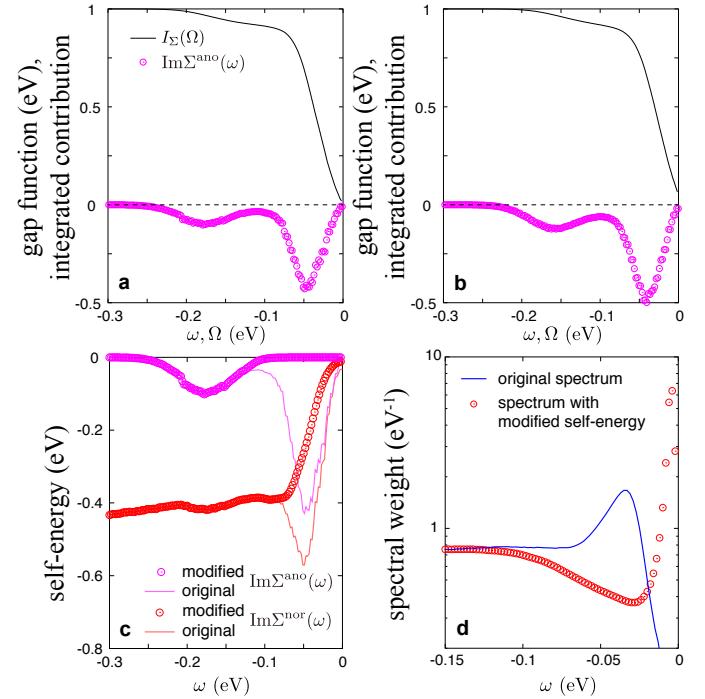


FIG. 3. Contribution of peak of $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ showing dominance to superconductivity. $I_{\Sigma}(\Omega)$ calculated from $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ is shown for the Bi2212 (**a**) and for the Bi2201 (**b**). The right negative peaks of $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ both contribute to more than 90% of $\text{Re}\Sigma^{\text{ano}}(\omega = 0)$. **c.** $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ modified from the originals in Figs. 2c. and 3a. by eliminating the low-energy peaks around $\omega = -50$ meV for Bi2212. The peak component of $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ to be subtracted is $\text{Im}\Sigma_{\text{PEAK}}(k_{\text{AN}}, \omega)$ in Fig. 2c and the subtracted $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ is nothing but $\text{Im}\Sigma_N(k_{\text{AN}}, \omega)$. On the other hand, $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ consists of only two peaks and the right peak around $\omega = -50$ meV can be easily subtracted by using the sigmoid function. Peak-subtracted $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ are represented by red and purple circles, respectively. (See Appendix A 5 (Decomposition of self-energy)) **d.** In comparison to the experimental $A(k, \omega)$ (blue thin curve), spectral function obtained from the peak-subtracted $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$ is shown by red circles, where superconductivity disappears resulting in a good normal metal with a quasiparticle width comparable to the experimental resolution (~ 10 meV).

To further ensure the crucial role of the inner peak in $\text{Im}\Sigma^{\text{ano}}(k_{\text{AN}}, \omega)$, we have hypothetically eliminated the peaks in $\text{Im}\Sigma^{\text{nor}}$ and $\text{Im}\Sigma^{\text{ano}}$ as shown in Fig. 3c. for Bi2212. The resultant $A(k, \omega)$ in Fig. 3d. shows the gap

disappearance and switching to a normal metal with a sharp quasiparticle peak, confirming the crucial role of the peaks to superconductivity.

Through our Boltzmann-machine analyses, Σ^{nor} , Σ^{ano} and W are revealed to have prominent (positive or negative) peaks, while they cancel in the sum Σ^{tot} . It is important that this conclusion is obtained directly from experimental data without assuming any theoretical models aside from mathematical requirement for the Green's function. A recent self-energy analysis of ARPES data¹³ did not identify the present prominent structure. The origin of the discrepancy is not perfectly clear but we discuss the origin in Appendix C.

If a large superconducting gap is open around $\omega = 0$ as in the experimental $A(k, \omega)$, it requires the corresponding famous gap structure in $\text{Re}\Sigma^{\text{ano}}$ around $\omega = 0$, where inside the two peaks at the gap energy $\omega = \pm\Delta$, $\text{Re}\Sigma^{\text{ano}}$ shows plateau. Then consistency requires a prominent peak structure around ω_{PEAK} in $\text{Im}\Sigma^{\text{ano}}$ through the Kramers-Kronig relation. The peak of $\text{Im}\Sigma^{\text{ano}}$ in turn naively anticipates prominent structures outside the gap in $A(k, \omega)$ through Eq.(1). However, such structures are missing. This is possible only when Σ^{nor} plays a role to cancel the prominent structure in $\text{Im}W$. This is corroborated by the vanishing superconductivity in Fig.3d.. Furthermore, the superconducting order accompanied by coherent quasiparticle excitations observed in experiments can be generated from $\text{Im}\Sigma^{\text{ano}}(\omega)$ only when electrons at ω become coherent, signaled by the reduction of $|\text{Im}\Sigma^{\text{tot}}|$ (or $|\text{Im}\Sigma_N|$) (see Figs.2c. and d., their captions and Appendix A) seen in the region $|\omega| < \omega^* \sim 0.07$ eV. This restricts ω_{PEAK} to this range. The present machine learning indeed reproduced this natural expectation in a physically reasonable way.

In Appendix D and E, we show some of analyses on temperature dependence including a case above T_c and momentum dependence away from the antinodal point, respectively as supporting data of the present analyses. The results confirm the validity of the present conclusion.

Although we found that the prominent peak in $\text{Im}\Sigma^{\text{ano}}$ is the true origin of the high critical temperature of the curates, full understanding and mechanism of prominent peaks in $\Sigma^{\text{nor}}(\omega)$ and $\Sigma^{\text{ano}}(\omega)$, which are absent in $\Sigma^{\text{tot}}(\omega)$ are open to further analyses. One possible explanation is the two-component fermion theory, where electrons are fractionalized into the bare electrons and dark fermions (hidden fermions) consistently with the cluster dynamical mean-field theory (cDMFT)¹⁶. If the spontaneous symmetry breaking such as the stripe order coexists with the superconductivity, the cancellation may be accounted for as an alternative interpretation^{16,17}.

IV. ROLE OF INTRINSIC (PLANCKIAN) DISSIPATION

The present machine learning is also useful to separately extract other theoretically fundamental quantities

such as the momentum resolved superconducting order parameter (the density of Cooper pair or the superfluid density) $F(k)$, mass renormalization factor $z_{\text{qp}}(k)$ and the single-particle relaxation time τ , which had been inferred only indirectly or only in combinations of more than one quantity in experiments in the literature, although these quantities play crucial roles below in understanding physics. (see Appendix A 1 for precise definition of the above quantities).

How frequently the single-particle excitations are scattered is encoded in the imaginary part of the normal self-energy $\text{Im}\Sigma^{\text{nor}}$. Landau's Fermi-liquid-like behavior characterized by $\text{Im}\Sigma^{\text{nor}}(\omega) \propto \omega^2$, is satisfied only in a small region ($|\omega| < 0.03$ eV) for Bi2212, and looks even linear ($\propto |\omega|$) in the same region for Bi2201, implying non-Fermi liquid (marginal Fermi liquid) behavior¹⁸ (see Figs. 2c. and d.), which can be fit by $\text{Im}\Sigma^{\text{nor}}(k, \omega) \sim c_0(k) + \text{sign}(\omega)c_1(k)\hbar\omega$ in the range $15 \text{ meV} < \omega < 40 \text{ meV}$ with a dimensionless marginal-Fermi-liquid coefficient $c_1(k)$. The ω -linear component $c_1(k)\omega$ is disruptive to the quasiparticle picture, and manifests emergent inelastic dissipation absent in Landau's Fermi liquids. As supporting information, tiny quasi-particle renormalization factor z_{qp} corroborating the non-Fermi liquid together with its effects on pair breaking is also shown in Appendix J (see definition of z_{qp} in Appendix A 1).

The single particle relaxation time τ is defined by $\tau(k, \omega)^{-1} = z_{\text{qp}}(k)\text{Im}\Sigma^{\text{nor}}(k, \omega)/\hbar$. When the carrier relaxation time is estimated from τ , the ω -linear term, $z_{\text{qp}}(k)c_1(k)\omega$, is associated with the universally observed T -linear resistivity in the cuprates^{19,20} through the ω - T correspondence $\tau(\hbar\omega) \leftrightarrow \tau(k_B T)$ transformed to the self-energy of two-particle Green's function for the conductivity. Appendix K showing temperature-insensitive $z_{\text{qp}}(k)c_1(k)$ also supports the correspondence. (See Fig. 17 of Appendix L for each $z_{\text{qp}}(k)$ and $c_1(k)$.)

A remarkable property of the inelastic relaxation rate $\Gamma(k) = z_{\text{qp}}(k)c_1(k)$ is its high value (~ 1) with only weak dependence on the doping, momentum (see Fig. 4a.) and temperature. This universal behavior of $\Gamma \sim 1$ -1.5 seems to support a local and universal mechanism of the relaxation, for instance, the Planckian dissipation mechanism of the hydrodynamic state, which claims $\tau^{-1}(T) = \Gamma k_B T/\hbar$ or $\tau^{-1}(\omega) = \Gamma\omega$ with a universal constant Γ of the order unity²¹⁻²³.

Although simple version of Planckian mechanism expects only an extended broad self-energy structure due to "unparticle physics," the self-energy has a broad but prominent peak structure around $\omega = \omega_{\text{PEAK}}$ which is responsible for the superconductivity through the Kramers-Kronig transformation as we discussed. At the same time, although the cDMFT studies¹⁶ suggest a sharp peak structure, the actual line shape is rather broad with the width around 0.05 eV (see Figs. 2c. and d.), which is comparable to ω_{PEAK} itself. More importantly, the peak is smoothly connected in the tail with the ω -linear behavior near the zero energy, implying that the "Planckian dissipation and hydrodynamic behavior" associated

with the strange metal²¹ is caused by the origin of the superconductivity. The broad prominent peaks could be due to the damped pole but it could also be ascribed to “unparticle object” generated by entangled bare electron and dark object. If the electron is hybridized with the fermionic part of excitonic excitations¹⁷, the strong entanglement of the exciton and electron can generate the unparticle feature, which may be ultimately ascribed to Mott physics through emergent excitonic bound states characteristic of (doped) Mott insulator. Alternatively, other mechanisms can be equally proposed. The present result poses severe constraints on possible theories.

V. RELATION BETWEEN THE SUPERCONDUCTING CRITICAL TEMPERATURE AND SUPERFLUID DENSITY

Fundamental quantities revealed by the machine learning provide further insight into the superconductivity through the scaling among experimental observables: The linear relation $F \propto T_c$ between T_c and the superfluid density F measured from the muon-spin relaxation rate R (theoretically proportional to $F(k)z_{qp}(k)$ averaged over Fermi surface momentum) has been examined through the Uemura plot²⁷ in high- T_c superconductors as in an example of the purple triangles in the inset of Fig. 4b. for Bi2201²⁴. The linearity should be satisfied for attractive interaction stronger than the effective Fermi energy scale E_F , which is proportional to the carrier density in two spatial dimensions. Here, E_F is roughly the effective bandwidth of the dispersion $z_{qp}\epsilon_k$. This can be regarded as the Bose-Einstein condensation (BEC) regime. Homes *et al.*²⁶ proposed empirical but more universal fitting as plotted in an example of Bi2201 by blue upside-down triangles in the inset of Fig. 4b.^{24,25}, where the dc conductivity σ_{dc} at T_c enters as $R \sim CT_c\sigma_{dc}(T_c)$ with a material independent constant C . Since σ_{dc} is proportional to the momentum relaxation time, the Homes relation proposes qualitatively different physics involving dissipation and scattering effects beyond the naive BEC regime. However, since σ_{dc} is believed to be proportional to both the carrier density and the relaxation time, it is not easy to single out the relaxation effect. Related scaling of the superfluid density Fz_{qp} proportional to the quasiparticle peak weight was also proposed^{28,29}.

Here, by replacing σ_{dc} by the microscopic relaxation rate $\Gamma(k)\omega$, we obtain a better linear scaling than the Uemura and Homes plots from the linear regression analysis, if we employ $F\bar{Q}\omega_0 \propto T_c(\Gamma(k))^{-1}$ with ω_0 taken as a doping-independent energy scale for Bi2201. Here, instead of z_{qp} defined in the $\omega \rightarrow 0$ limit, in the left hand side, we employ the renormalization factor Q at k_{AN} averaged in the self-energy peak region, namely,

$$\bar{Q} = \int d\omega \text{Im}W_{\text{PEAK}}(k, \omega) Q(k, \omega) / \overline{W_{\text{PEAK}}}(k) \quad (5)$$

with the integration over the interval $\omega < 0$, where the integrated peak intensity of the self-energy peak $\text{Im}\Sigma_{\text{PEAK}} = -\text{Im}W_{\text{PEAK}}$ plotted as the yellow area in Figs. 2c. and d. is defined as $\overline{W_{\text{PEAK}}}(k) = \int d\omega \text{Im}W_{\text{PEAK}}(k, \omega)$. The replacement of z_{qp} with \bar{Q} allows to include the pairing energy scale reduced by the pair breaking effect \bar{Q} more appropriately and it indeed gives better fitting. However, material independent ω_0 does not explain the large T_c difference between Bi2201 and Bi2212, because $\Gamma(k)$ and $F\bar{Q}$ are similar at the optimum doping (see Table I).

Material-dependent T_c can be represented by

$$\Omega_0 = \frac{\overline{W_{\text{PEAK}}}(k_{AN})}{\omega_{\text{PEAK}}(k_{AN})}, \quad (6)$$

instead of a constant ω_0 . Note that Ω_0 is the correct scale of superconducting amplitude because Ω_0 measures the contribution to the superconducting gap through the Kramers-Kronig relation, where the peak of $\text{Im}\Sigma_{\text{PEAK}}$ (yellow area in Figs. 2c. and d.) and $\text{Im}W_{\text{PEAK}}$ (pink area in Figs. 2c. and d.) have the same amplitude, and hence $\text{Im}\Sigma^{\text{ano}}$ does have a dominant peak as well. In fact, the primary origin of the large difference of T_c between Bi2201 and Bi2212 is identified as the difference in $\overline{W_{\text{PEAK}}}$ (namely the coupling strength of the electron with the dark object which makes the prominent self-energy peaks), supplemented by the difference in ω_{PEAK} . (see Table I. $\overline{W_{\text{PEAK}}}(k_{AN}) \sim 7.6 \times 10^{-3}$ eV² and $\omega_{\text{PEAK}}(k_{AN}) \sim 0.07$ eV for the optimum Bi2201 and $\overline{W_{\text{PEAK}}}(k_{AN}) \sim 1.4 \times 10^{-2}$ eV² and $\omega_{\text{PEAK}}(k_{AN}) \sim 0.045$ eV for the optimal Bi2212 at $k = k_{AN}$). (See also the list of Ω_0 in Table I and the angle and doping dependences of $\overline{W_{\text{PEAK}}}(k)/\omega_{\text{PEAK}}(k)$, and $\bar{Q}(k)$ in Fig. 17 of Appendix L.)

Now, as shown in Figs. 4b.,

$$k_B T_c = \Gamma(k_N) g(k_{AN}) F(k_{AN}) \quad (7)$$

with

$$g(k_{AN}) = \bar{Q}(k_{AN}) \Omega_0(k_{AN}) \quad (8)$$

gives the best linear fit (the main panel of Fig. 4b.). The transition temperature T_c is determined first by F of course, but it is also scaled by the effective Cooper-pair-formation energy scale of the attractive coupling given by g primarily around the antinodal region and the degree of quantum entanglement among, for instance, the bare electron and the emergent dark object together with the entanglement within the quasiparticles measured by $\Gamma(k_N)$ in the nodal region.

Although the Homes plot does not offer how T_c is determined because $\sigma(T_c) \propto 1/T_c$ cancels in the relation to F , the present result indeed shows T_c linearly scaled by $\Gamma(k_N) g(k_{AN}) F(k_{AN})$. The linearity is crucially different from the Uemura plot as well because of the dependence on the relaxation rate Γ . Intuitively, $\tau = \hbar/(\Gamma k_B T)$ or $\hbar/(\Gamma k_B \omega)$ is related to the characteristic length scale λ for the extension of the quantum mechanically entangled

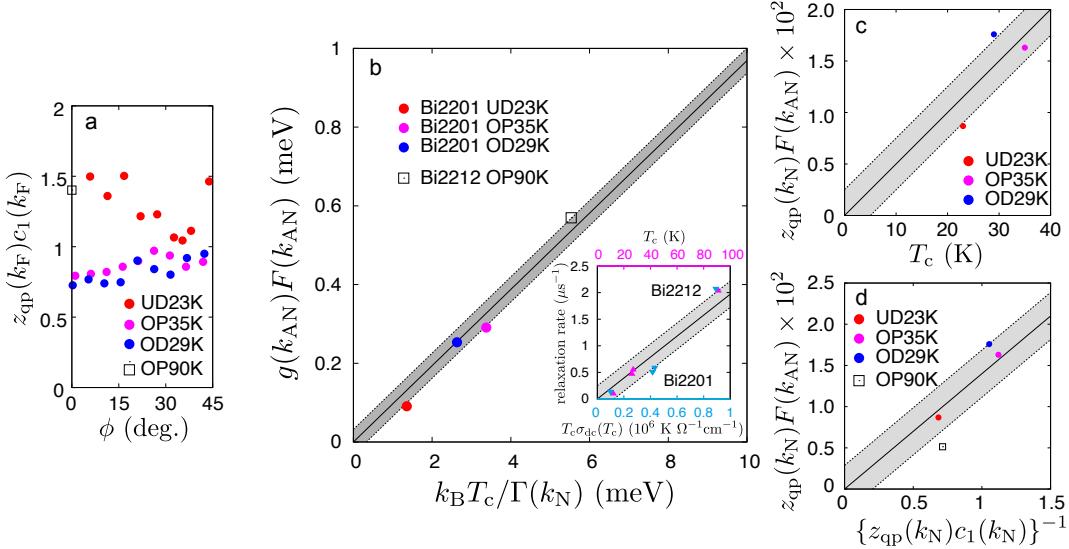


FIG. 4. Relation between superfluid density F , T_c , carrier relaxation time and self-energy peak derived from machine learning. **a.** Angle and doping dependences of $\Gamma = z_{\text{qp}} c_1$ for Bi2201 and Bi2212. **b.** Proposed scaling between F , T_c , g , and Γ . **c.** Possible scaling between F and T_c for Bi2201, which mimics the Uemura plot (see text). **d.** Possible scaling between F and Γ^{-1} for Bi2201 and Bi2212, which mimics the Homes plot (see text). The scaling $gF \propto T_c/\Gamma$ in **b.** gives the best fitting: The standard deviation is 0.03 meV for **b.**, while 0.25 for **c.** (excluding the 2212 data because it is far away from the linear fitting) and 0.29 for **d.**. Even for the standard deviation of normalized values, $(a_i - a_i^{\text{fit}})/a_i^{\text{fit}}$ with a_i (a_i^{fit}) being the i -th data of $\bar{Q}F$ or $z_{\text{qp}}F$ (the fitting line value), instead of $a_i - a_i^{\text{fit}}$ employed above, we obtain 0.023 (**b.**), 0.19 (**c.** excluding 2212) and 0.26 (**d.**). Inset of **b.**: Experimental plots of the muon-spin relaxation rate R ²⁴ vs. T_c (Uemura plot) or $T_c\sigma_{\text{dc}}(T_c)$ (Homes plot)^{25,26} for Bi2201 and Bi2212. Here, the standard deviation is $0.07 \mu\text{s}^{-1}$ ($0.25 \mu\text{s}^{-1}$), and the standard deviation of the normalized values is 0.22 (0.26) for the Uemura plot (Homes plot).

area through $\lambda \sim v_F \tau^{21}$, where v_F is the characteristic electron velocity (“Fermi velocity”). The larger attraction generates stronger self-energy peak. It necessarily generates the steeper ω -linear tail of $\text{Im}\Sigma^{\text{nor}}$ near zero energy, further enhancing more local and stronger pairing approaching the BEC limit, and raises T_c through Eq.(7). It seems that the strange metal and high T_c with strong attraction, in this sense, represent the two sides of the same coin. It is also interesting to note that Eq.(7) looks compatible with the scaling $E_c \propto \gamma_c T_c^{230}$, where E_c is the condensation energy and γ_c is the Sommerfeld constant of the specific heat, because gF plays the role of the gap, which generates the energy gain.

Note that $\Gamma(k)$ should be analyzed around k_N , while Ω_0 and F contribute at k_{AN} for better fitting. The in-plane transport and the quantum entanglement are dominated by the contribution around the nodal region, while the pairing looks driven in the antinodal region.

Equation (7) is the best scaling among various attempts we have made. To convince readers, we just show two examples of plot in Figs. 4c. and d.. The first example is $z_{\text{qp}}(k_{\text{AN}})F(k_{\text{AN}})$ vs. T_c plotted in Fig. 4c., which, though not perfectly equivalent, apparently mimics the Uemura plot. The second is $z_{\text{qp}}(k_N)F(k_{\text{AN}})$ vs. $1/z_{\text{qp}}(k_N)c_1(k_N)$ which mimics the Homes plot, because $T_c\sigma(T_c) \propto 1/z_{\text{qp}}(k_N)c_1(k_N)$ is expected (see Fig. 4d.).

The standard deviation is by far best for the present fit in Fig. 4b. with Eq.(7).

The present result is significant because the whole analyses are obtained solely from the single ARPES line shape of $A(k, \omega)$ and contains much less ambiguity than before. The present machine learning purely from experimental data sheds new light on understanding the superconducting mechanism, where the energy dissipation plays a role through the extension of quantum entanglement. For detailed doping concentration and momentum dependences of T_c , $F(k)$, $c_1(k)$, $z_{\text{qp}}(k)$, and the superconducting gap $\Delta_0(k)$ for Bi2201 at 11K are found in Appendix L.

VI. OUTLOOK

It is desired to examine the present results on the cancellation of two self-energy contributions which makes the superconducting temperature high and the role of dissipation in determining T_c in other cuprate compounds by measuring $A(k, \omega)$ at high accuracy and resolution. Present successful examples of insight obtained purely from the machine learning analysis of experimental data indicates an opening of a promising field which allows understanding physics hidden in experiments, without

relying on involved and specific theoretical assumptions and biases. By combining with other experimental data and indisputable theoretically basic constraints such as symmetry, much more powerful tool will be provided for understanding physics of complex phenomena.

ACKNOWLEDGMENTS

We thank Takeshi Kondo and Adam Kaminski for providing us his ARPES data published in Refs. 9 and 10. We also thank Takeshi Kondo for discussions on the experimental results. We are grateful to Shiro Sakai for discussions and comments on the manuscript. Y. Y. was supported by PRESTO, JST (JPMJPR15NF). Y. Y. and M.I. were supported by JSPS KAKENHI (Grant No. 16H06345). This research was also supported by MEXT as “Priority Issue on Post-K computer” (Creation of New Functional Devices and High-Performance Materials to Support Next-Generation Industries (CDMSI)) with the project supported by RIKEN Advanced Institute for Computational Science (AICS) through HPCI System Research Project (Grants No. No. hp170263 and hp180170).

Appendix A: Methods

The present machine learning scheme is classified into the category of a general regression task, which optimizes a function A to find B and/or F , when A is a non-linear and complex functional of another function B as $A = F(B)$, where a goal of optimization of A is set from a physical purpose³¹. Another example belonging to the regression task as an application of the machine learning is found in the use for quantum many body problem or statistical physics problem (see for instance, Ref. 32). In the present case, A is the ARPES spectral function $A(k, \omega)$, B is the self-energies Σ^{nor} and Σ^{ano} and F is given from Eqs. (1) and (2). The training data is given by experimental $A(k, \omega)$ at a discrete and limited number of ω . Then the present machine learning is a typical regression task to infer $\Sigma^{\text{ano}}(k, \omega)$ and $\Sigma^{\text{nor}}(k, \omega)$ separately as continuous functions of ω . The basic procedure is illustrated in Figure 5. The robustness, accuracy and reliability of the present machine learning are shown in Appendix B, C and F.

1. Green’s function

We propose a theoretical method to extract Σ^{nor} and Σ^{ano} from experimentally observed spectral functions $A(k, \omega)$ of superconductors. In superconducting phases, the single-particle retarded Green’s function at a given momentum k as a function of frequency ω is given by a diagonal component of 2×2 matrix in Nambu representation,

tation,

$$\hat{G}(k, \zeta) = \begin{bmatrix} \zeta - \epsilon_k - \Sigma^{\text{nor}}(k, \zeta) & -\Sigma^{\text{ano}}(k, \zeta) \\ -\Sigma^{\text{ano}}(k, \zeta) & \zeta + \epsilon_k + \Sigma^{\text{nor}}(k, -\zeta)^* \end{bmatrix}^{-1}, \quad (\text{A1})$$

with $\zeta = \omega + i\delta$ (δ is a small positive real number). The bare dispersion is given by ϵ_k . $A(k, \omega)$ measurable by ARPES is related to \hat{G} as

$$A(k, \omega) = -\frac{1}{\pi} \text{Im} [\{\hat{G}(k, \zeta)\}_{11}]_{\delta \rightarrow +0}, \quad (\text{A2})$$

with the normal component of the Green function

$$G^{\text{nor}}(k, \omega) \equiv \hat{G}(k, \omega)_{11} = [\omega - \epsilon_k - \Sigma^{\text{tot}}(k, \omega)]^{-1}, \quad (\text{A3})$$

while the total self-energy Σ^{tot} is given by¹²

$$\Sigma^{\text{tot}}(k, \omega) = [\Sigma^{\text{nor}}(k, \zeta) + W(k, \zeta)]_{\delta \rightarrow +0}. \quad (\text{A4})$$

with W given in Eq. (2) In the gap function Eq. (3) the renormalization is defined as

$$Q(k, \omega) = \left. \frac{1}{1 - [\Sigma^{\text{nor}}(k, \zeta) - \Sigma^{\text{nor}}(k, -\zeta)^*]/(2\zeta)} \right|_{\delta \rightarrow +0}. \quad (\text{A5})$$

The $\omega \rightarrow 0$ limit of Q is theoretically equivalent to the quasiparticle weight (renormalization factor) defined in Eq.(A7) below as we calculate in Appendix J. The real part of $\Delta(k, \omega)$, $\text{Re}\Delta(k, \omega = 0)$ is nothing but the superconducting gap (see the definition of Δ in Eq.(3)). In the present paper, δ is chosen to be equal to the experimental resolution as $\delta = 10 \text{ meV}^9$, instead of taking $\delta \rightarrow 0^+$.

To estimate the density of the Cooper pairs, mass renormalization, and gap amplitude in the spectral weight, we define $F(k)$, $z_{\text{qp}}(k)$, and $\Delta_0(k)$ respectively as,

$$F(k) = \int_{-\infty}^0 d\omega \frac{1}{\pi} \text{Im} \hat{G}(k, \omega)_{12}, \quad (\text{A6})$$

$$z_{\text{qp}}^{-1}(k) = 1 - \partial \text{Re} \Sigma^{\text{nor}}(k, \omega) / \partial \omega |_{\omega \rightarrow 0}, \quad (\text{A7})$$

$$\Delta_0(k) = \Delta(k, \omega = \Delta_0(k)), \quad (\text{A8})$$

where $\Delta(k, \omega)$ is defined in Eq.(3) and z_{qp} is called the renormalization factor (see Appendix J).

2. Boltzmann machine representation of imaginary part of self-energy

Although high-resolution ARPES data for $A(k, \omega)$ are available in experiments, Σ^{nor} and Σ^{ano} are not directly given separately from $A(k, \omega)$, while if Σ^{nor} and Σ^{ano} and ϵ_k are given, $A(k, \omega)$ can be determined easily by using Eqs. (A1) and (A2). Therefore we need to solve an underdetermined non-linear inverse problem. To overcome the underdetermined nature of the problem, we employ physical constraints as prior knowledge as detailed below. By incorporating these physical constraints, we try to

optimize Σ^{nor} and Σ^{ano} so as to reproduce experimental $A(k, \omega)$. For this purpose we employ a machine-learning method by applying a Boltzmann-machine algorithm⁷. The reliability, accuracy and robustness of the present machine learning procedure are shown in several robustness test against noise in Appendix B and benchmark tests in Appendix F.

The physical constraints employed in the present paper are classified into two categories. The constraints in the first category are the structure of the Green's function given in Eq.(A1), the Kramers-Kronig relationship between the real and imaginary part of the self-energies, negative definiteness of $\text{Im}\Sigma^{\text{nor}}$, and odd nature of $\text{Im}\Sigma^{\text{ano}}$ as $\text{Im}\Sigma^{\text{ano}}(-\omega) = -\text{Im}\Sigma^{\text{ano}}(\omega)$. The constraint in the second category is the sparse and localized nature of the $\text{Im}\Sigma^{\text{ano}}$ along ω axis. In the present context, the sparseness is defined as the property of $\text{Im}\Sigma^{\text{ano}}$ concentrated and localized in the small frequency range around the Fermi level. It should be mentioned that the optimization procedure of the present machine learning does not explicitly impose the constraint because of the flexible representability of the present (restricted) Boltzmann machine³³. The sparse nature of $\text{Im}\Sigma^{\text{ano}}$ is, however, justified *a posteriori* eventually in the optimized solution, which turns out to satisfy physically reasonable sparseness though the machine learning procedure does not explicitly impose this constraint.

From physical grounds, the sparse and localized nature of $\text{Im}\Sigma^{\text{ano}}$ is a natural consequence as clarified in Ref. 34, where irrespective of the mechanism and symmetry of the realistic pairing, strong and long-range nature of Coulomb repulsion in general causes severe pair breaking at large energies and suppresses Σ^{ano} at energies far away from the fermi level.

In the retarded Green's function representation, $\text{Im}\Sigma^{\text{nor}}$ is negative definite. Then $\text{Im}\Sigma^{\text{nor}}$ can be fit by

$$\text{Im}\Sigma^{\text{nor}}(\omega) = - \sum_{\sigma} C(\sigma) \Theta_{\sigma}^L \left(\frac{\omega + \Lambda/2}{\Lambda} \right), \quad (\text{A9})$$

where C is represented by a superposition of a series of rectangular functions $\Theta_{\sigma}^L(x)$ defined in Eq.(A12) below. In this fitting, the frequency range of our interest $\omega \in [-\Lambda/2, +\Lambda/2]$ is first divided into 2^L grids using an integer L and assign an L -digit binary representation as

$$\sigma \equiv (\sigma_1, \sigma_2, \dots, \sigma_L), \quad (\text{A10})$$

where $\sigma_i = \text{mod}(I/2^{i-1}, 2)$ for the decimal representation $I(\sigma)$ in the range $1 \leq I(\sigma) \leq 2^L$ of the grid number coordinate;

$$I(\sigma) = \sum_{\ell=0}^{L-1} \sigma_{\ell} \cdot 2^{\ell}. \quad (\text{A11})$$

The grid mesh is chosen to be smaller than or comparable to the experimental energy resolution ($\sim 10\text{meV}$ ^{9,10}) to fully reproduce the experimental $A(k, \omega)$ within the

resolution of the grid size $\Delta\omega = \Lambda/2^L$. Then the unit rectangular function $\Theta_{\sigma}^L(x)$ is defined as

$$\begin{aligned} \Theta_{\sigma}^L(\omega) &= 1 \quad \text{for } x \in [I(\sigma)/2^L, \{1 + I(\sigma)\}/2^L], \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (\text{A12})$$

The nonnegative weight $C(\sigma)$ is to be determined by the optimization using the Boltzmann machine later.

Thanks to the non-negativity of $C(\sigma)$, it is efficiently optimized by the Boltzmann machine. The Boltzmann machine is generally defined as a Boltzmann weight for Ising variables $\nu_{\ell} = \pm 1$

$$\mathcal{B}(\nu | \mathbf{W}, \mathbf{B}) = \exp \left[\sum_{\ell, m} W_{\ell m} \nu_{\ell} \nu_m + \sum_{\ell} B_{\ell} \nu_{\ell} \right], \quad (\text{A13})$$

where $(\mathbf{W})_{\ell m} = W_{\ell m}$ represents interaction among ν , and $(\mathbf{B})_{\ell} = B_{\ell}$ represents magnetic fields applied to ν .

The restricted Boltzmann machine (RBM)^{8,35,36} is one of the most widely used Boltzmann machine and introduces hidden Ising variables \mathbf{h} in addition to the physical Ising variables \mathbf{S} that constitute the whole Ising variables as $\nu = (\mathbf{S}, \mathbf{h})$. The RBM allows the interactions only between *visible* and *hidden* variables in the form $W_{\ell m} S_{\ell} h_m$. While the visible variables correspond to physical degrees of freedom, the hidden variable is internal one. Here, we assign the visible Ising variables $\mathbf{S} = (S_1, S_2, \dots, S_{2^L})$ for σ introduced in Eq.(A10) using the relation $S_{\ell} = 2\sigma_{\ell} - 1$. With the hidden variables \mathbf{h} , a flexible Boltzmann machines including the RBM is constructed as

$$\mathcal{C}(\mathbf{S}) = \sum_{\mathbf{h}} \mathcal{B}(\mathbf{S}, \mathbf{h} | \mathbf{W}, \mathbf{B}), \quad (\text{A14})$$

where we rewrite $C(\sigma)$ as $\mathcal{C}(\mathbf{S}) = C(\sigma)$ by changing the variable from the binary variable σ to the Ising variable \mathbf{S} . We can further generalize the Boltzmann machine by making a mixture distribution of the Boltzmann machine with coefficients w_{λ} as

$$\mathcal{D}(\mathbf{S}) = \sum_{\lambda} \sum_{\mathbf{h}} w_{\lambda} \mathcal{B}(\mathbf{S}, \mathbf{h} | \mathbf{W}^{\lambda}, \mathbf{B}^{\lambda}). \quad (\text{A15})$$

In this paper, we employ two different types of Boltzmann machines to enhance their representational power and determine each Σ^{nor} and Σ^{ano} , separately. $\text{Im}\Sigma^{\text{nor}}$ is negative definite and a widely distributed function within energy scale set by the Coulomb repulsion. Thus, we use the flexible and nonnegative RBM. However, $\text{Im}\Sigma^{\text{ano}}$ has different properties: It is sparse, which is justified *a posteriori* as we addressed already. Therefore, we employ a mixture distribution of the Boltzmann machine Eq.(A15) without the hidden variables to accelerate the optimization. There may still remain ambiguities in determining Σ^{nor} and Σ^{ano} from an observed $A(k, \omega)$. As proposed below, the ambiguities are removed by imposing physical constraints of $\text{Im}\Sigma^{\text{ano}}$.

Our restricted Boltzmann machine for $-\text{Im}\Sigma^{\text{nor}}$ is given by

$$\mathcal{C}(\mathbf{S}) = e^b \sum_{\{h_m=\pm 1\}} e^{\sum_{\ell,m} S_\ell W_{\ell m} h_m}, \quad (\text{A16})$$

in the expression Eq.(A9), where $W_{\ell m}$ and b are variational parameters to minimize the difference between $A(k, \omega)$ obtained from Eq.(A2) to (A4) and the experimental value.

The advantage of the RBM is that one can analytically trace out the hidden variables h_m , leading to

$$\mathcal{C}(\mathbf{S}) = e^b \prod_{m=0}^{M-1} 2 \cosh [S_\ell W_{\ell m}]. \quad (\text{A17})$$

Any ω -dependent line shape in the energy range $[-\Lambda/2, \Lambda/2]$ can be flexibly represented by optimized Boltzmann-machine parameters, if they are nonnegative.

For Σ^{ano} , to remove the ambiguities, we impose the physically required symmetries,

$$\begin{aligned} \text{Re}\Sigma^{\text{ano}}(\omega) &= \text{Re}\Sigma^{\text{ano}}(-\omega) \\ \text{Im}\Sigma^{\text{ano}}(\omega) &= -\text{Im}\Sigma^{\text{ano}}(-\omega), \end{aligned} \quad (\text{A18})$$

which can be constrained by employing a mixture distribution. Namely, we represent Σ^{ano} as a mixture distri-

bution consisting of Boltzmann machines as

$$\begin{aligned} \text{Im}\Sigma^{\text{ano}}(\omega) &= \sum_{\boldsymbol{\sigma}} D(\boldsymbol{\sigma}) \left[\Theta_{\boldsymbol{\sigma}}^L \left(\frac{\omega + \Lambda/2}{\Lambda} \right) \right. \\ &\quad \left. - \Theta_{\boldsymbol{\sigma}}^L \left(\frac{\Lambda/2 - \omega}{\Lambda} \right) \right], \end{aligned} \quad (\text{A19})$$

If it is sparse, Σ^{ano} can be better represented by Eq.(A20) below, as we see in an example of the Gaussian distribution in Appendix G.

$$\mathcal{D}(\mathbf{S}) = \sum_{\lambda=1}^M w_{\lambda} e^{\sum_{\ell,m} S_\ell V_{\ell m}^{\lambda} S_m + \sum_{\ell} S_\ell b_{\ell}^{\lambda}} \quad (\text{A20})$$

is the weight for the anomalous part represented by the Boltzmann machine and w_{λ} , $V_{\ell m}^{\lambda}$ and b_{ℓ}^{λ} are variational parameters to minimize the difference between the resultant $A(k, \omega)$ and the measured spectral functions.

Since it is sufficient to take the number of variables S_ℓ at most 9 to fit the experiment data containing the resolution limitation, we can explicitly take the trace summation over S_ℓ for all ℓ (with 2^L terms) at each iteration step. Therefore the drawback of the form Eq.(A20) beyond the RBM (namely, the complexity arising from containing the term proportional to $V_{\ell m}^{\lambda}$) is not a serious problem.

3. Real part of self-energy

The real part of the retarded self-energy is obtained through the Kramers-Kronig relation as,

$$\text{Re}\Sigma^{\text{nor}}(\vec{k}, \omega) = \frac{1}{\pi} \mathcal{P} \int d\omega' \frac{\text{Im}\Sigma^{\text{nor}}(\vec{k}, \omega')}{\omega' - \omega}, \quad (\text{A21})$$

$$\text{Re}\Sigma^{\text{ano}}(\vec{k}, \omega) = \frac{1}{\pi} \mathcal{P} \int d\omega' \frac{\text{Im}\Sigma^{\text{ano}}(\vec{k}, \omega')}{\omega' - \omega}. \quad (\text{A22})$$

For example, the real part of normal self-energy is obtained as

$$\text{Re}\Sigma^{\text{nor}}(\omega) = - \sum_{\mathbf{S}} \frac{\mathcal{C}(\mathbf{S})}{2\pi} \ln \frac{\{\Lambda(1 + I(\boldsymbol{\sigma}))/2^L - \Lambda/2 - \omega\}^2 + \delta^2}{\{\Lambda I(\boldsymbol{\sigma})/2^L - \Lambda/2 - \omega\}^2 + \delta^2}, \quad (\text{A23})$$

where we introduce a broadening factor δ .

4. Numerical procedure for optimization

We optimize the Boltzmann machine to reproduce experimentally observed spectral weights. Naively, the least square error defined by

$$\chi^2 = \frac{1}{2N_d} \sum_j \{A^{\text{exp}}(\omega_j) - f(\omega_j)A(\omega_j)\}^2, \quad (\text{A24})$$

is minimized, where N_d is the number of the experimental data points, $A^{\text{exp}}(\omega)$ is an experimentally observed $A(k, \omega)$, $\{\omega_j\}$ ($j = 1, 2, \dots, N_d$) is the set of frequency

where $A(k, \omega)$ is observed in the experiment, and $f(\omega_j)$ is a convolution of the Fermi-Dirac distribution and a Gaussian distribution. The experimental data $A^{\text{exp}}(\omega_j)$ involves the Fermi-Dirac distribution broadened by the resolution of the experiments. Therefore, we introduce the convolution $f(\omega)$ of the Fermi-Dirac distribution at 12 K for Bi2212 and 11 K for Bi2201 and the Gaussian distribution with standard deviation $\sqrt{\sigma^2} = 5$ meV. Here, we normalize the experimental data $A^{\text{exp}}(\omega)$ by assuming $(1/N_d) \sum_j A^{\text{exp}}(\omega_j) = n_0$, where $0 < n_0 < 1$. In this paper, we infer $n_0 = 0.3$ per spin, which means

that 60% of an electron is assumed to be distributed for $\omega \gtrsim -0.4$ eV. We show that $n_0 = 0.3$ is indeed the optimized value of the least square fit in Appendix H while the result of the self-energies does not sensitively depend on the choice of n_0 around 0.3.

The parameters in the Boltzmann machine, $\boldsymbol{\alpha}^{\text{nor}} = (b, \{W_{\ell m}\})$ and $\boldsymbol{\alpha}^{\text{ano}} = (\omega_\lambda, \{b_\ell^\lambda\}, \{V_{\ell m}^\lambda\})$, are optimized by using the standard gradient method. The parameters at the k th step, $\boldsymbol{\alpha}_k^{\text{nor}}$ and $\boldsymbol{\alpha}_k^{\text{ano}}$, are updated as

$$\boldsymbol{\alpha}_{k+1}^{\text{nor}} = \boldsymbol{\alpha}_k^{\text{nor}} - \epsilon (\|S^{-1}\mathbf{g}_k^{\text{nor}}\|_1)^{-1/2} S^{-1}\mathbf{g}_k^{\text{nor}}, \quad (\text{A25})$$

$$\boldsymbol{\alpha}_{k+1}^{\text{ano}} = \boldsymbol{\alpha}_k^{\text{ano}} - \epsilon' (\|\mathbf{g}_k^{\text{ano}}\|_1)^{-1/2} \mathbf{g}_k^{\text{ano}}, \quad (\text{A26})$$

where

$$S_{\mu\nu} = \frac{1}{N_d} \sum_j \frac{\partial \text{Im}\Sigma^{\text{nor}}(\omega_j)}{\partial \alpha_\mu^{\text{nor}}} \frac{\partial \text{Im}\Sigma^{\text{nor}}(\omega_j)}{\partial \alpha_\nu^{\text{nor}}}, \quad (\text{A27})$$

$$\mathbf{g}_k^{\text{nor}} = \frac{\partial \chi^2}{\partial \boldsymbol{\alpha}^{\text{nor}}}, \quad (\text{A28})$$

$$\mathbf{g}_k^{\text{ano}} = \frac{\partial \chi^2}{\partial \boldsymbol{\alpha}^{\text{ano}}}, \quad (\text{A29})$$

and $\|\cdots\|_1$ represents L_1 norm. The factors $(\|S^{-1}\mathbf{g}_k^{\text{nor}}\|_1)^{-1/2}$ and $(\|\mathbf{g}_k^{\text{ano}}\|_1)^{-1/2}$ are introduced to accelerate the optimization. Here, we use the natural gradient method to optimize the variational parameters in $\text{Im}\Sigma^{\text{nor}}(\omega_j)$ because the optimization is efficient,^{36–38} while the simple steepest descent method is employed to optimize the part of $\text{Im}\Sigma^{\text{ano}}(\omega_j)$ because the natural gradient method assumes that the optimized distribution is positive or negative definite. During the optimization of the Boltzmann machine, we may introduce a regularization term by L_1 norm of the mixture of the Boltzmann machines as $\lambda_w \sum_\lambda |w_\lambda|$. While $\lambda_w = 10^{-3}$ will accelerate the optimization, the results of the optimization is confirmed to be insensitive to $\lambda_w \leq 10^{-3}$. In the actual fitting, we employed $\lambda_w = 10^{-3}$.

To explore the multi-dimensional parameter space of the Boltzmann machine and find an optimized solution, we employ the Bayesian optimization scheme. First, we perform sufficiently large number of minimization steps, which is typically 4×10^3 , with an initial parameter set.

Then, we update the center of mass of each Boltzmann machine in $\text{Im}\Sigma^{\text{ano}}$ defined in Eq.(A19). To generate the next proposal for the center of mass, we use the Bayesian optimization scheme³⁹ depending on the history of the optimization process for the center of mass of each Boltzmann machine in $\text{Im}\Sigma^{\text{ano}}$. When we generate the proposal, we choose the proposal to minimize another cost function $\overline{\chi^2}$ defined below.

To avoid overfitting, we use $\overline{\chi^2}$ defined below to make the proposal instead of the χ^2 . First, we generate fictitious experimental data from the original data. Because the overfitting originates from reproducing detailed noisy behaviors in the experimental data finer than the experimental resolution⁴⁰, to eliminate the noise, the experimental data A^{exp} is fitted by a smooth function A^{fit} defined as a linear combination of the Gaussian distributions³⁹ with standard deviation $\sqrt{\sigma^2} =$

10 meV equal to the experimental resolution. Then, we can estimate amplitude of noise in the experimental data as $\sigma_n^2 = N_d^{-1} \sum_j (A^{\text{fit}}(\omega_j) - A^{\text{exp}}(\omega_j))^2$. By using the probability distribution $p(A'|A^{\text{fit}}, \omega) \propto \exp \left[- (A'(\omega) - A^{\text{fit}}(\omega))^2 / 2\sigma_n^2 \right]$, we can generate fictitious experimental data. If we assume that $p(A'|A^{\text{fit}}, \omega)$ well reproduces real experimental data, the cost function to avoid overfitting is defined by,

$$\overline{\chi^2} = \frac{1}{N_d N_r} \sum_{s=1}^{N_d} \sum_{r=1}^{N_r} \{A'_r(\omega_s^{(r)}) - A(\omega_s^{(r)})\}^2 f(\omega_s^{(r)})^2, \quad (\text{A30})$$

where A'_r is the r th fictitious experimental data generated by the probability distribution $p(A'|A^{\text{fit}}, \omega)$ and $\{\omega_s^{(r)}\}$ is a set of randomly chosen frequency points for each fictitious data A'_r .

The sequential optimization of the internal parameters of the Boltzmann machines and the center of mass of the mixture distribution is repeated several tens. The self-energies that give the minimum $\overline{\chi^2}$ during the process are shown as the inferred self-energies.

In the present paper, first, we optimize the Boltzmann machine with $L = 8$ visible nodes and $2L = 16$ hidden nodes for the part $\text{Im}\Sigma^{\text{nor}}$ and, then, we enhance the resolution with $L = 9$ visible nodes and 18 hidden nodes to obtain better resolution with reasonable numerical cost. In the optimization with $L = 9$, we avoid the update of the center of mass by the Bayesian process and perform longer minimization steps up to 2×10^4 . We employ the broadening factor $\delta = 10$ meV throughout this paper. We show in Appendix I that the result does not sensitively depend on the choice of δ .

5. Decomposition of self-energy

While the normal and superconducting components of the total self-energy, $\text{Im}\Sigma(k, \omega)$ and $\text{Im}W(k, \omega)$, show the prominent peak structures, which are absent in the Bardeen-Cooper-Schrieffer (BCS) mean-field theory⁴¹, there are an extended background and a BCS-like superconducting contribution, in addition to the peaks. To highlight these peaks, we decompose $\text{Im}\Sigma^{\text{nor}}(k, \omega)$ and $\text{Im}W(k, \omega)$ into the peaks and other components. As proposed in Ref. 14, $\Sigma^{\text{tot}}(k, \omega)$ may consist of a single pole that generates a superconducting gap and a smooth normal state component. Then, we decompose $\text{Im}\Sigma^{\text{tot}}(k, \omega)$ as

$$\text{Im}\Sigma^{\text{tot}}(k, \omega) = \text{Im}\Sigma_N(k, \omega) + L_{\text{BCS}}(k, \omega). \quad (\text{A31})$$

Here, while the BCS-like superconducting contribution is represented by a Lorentzian,

$$L_{\text{BCS}}(k, \omega) = -\frac{1}{\pi} \frac{\Delta_0^2 \Gamma}{(\omega + \epsilon_k)^2 + \Gamma^2}, \quad (\text{A32})$$

where Δ_0 and Γ are phenomenological parameters that correspond to a BCS-like superconducting gap and life time of quasiparticles, respectively. The background $\text{Im}\Sigma_N(k, \omega)$ is represented by a linear combination of many Gaussian distributions, where its large amplitude signals the incoherence of electrons at that energy. Then, the peak contribution cancelled in $\text{Im}\Sigma^{\text{tot}}(k, \omega)$ is obtained from $\text{Im}\Sigma^{\text{nor}}(k, \omega)$ as

$$\text{Im}\Sigma_{\text{PEAK}}(k, \omega) = \text{Im}\Sigma^{\text{nor}}(k, \omega) - \text{Im}\Sigma_N(k, \omega), \quad (\text{A33})$$

and from $\text{Im}W(k, \omega)$ as

$$\text{Im}W_{\text{PEAK}}(k, \omega) = \text{Im}W(k, \omega) - L_{\text{BCS}}(k, \omega), \quad (\text{A34})$$

where $\text{Im}\Sigma_{\text{PEAK}}(k, \omega) = -\text{Im}W_{\text{PEAK}}(k, \omega)$ holds.

6. Resolution of gap functions

The gap function $\Delta(\omega)$ defined in Eq. (3) can show significant δ dependence near the small δ limit around $\omega \sim 0$. The δ dependence originates from the finite imaginary part of the normal self-energy $\text{Im}\Sigma^{\text{nor}}(k, \omega = 0)$ inevitable in the experimental data. When we modify Q as

$$Q(k, \omega) = \frac{1}{1 - \frac{\Sigma^{\text{nor}}(k, \omega + i\delta) - \Sigma^{\text{nor}}(k, -\omega - i\delta)^*}{2(\omega + i\delta')}}, \quad (\text{A35})$$

we obtain stable behaviors of $\Delta(\omega)$ for $|\omega| > 10$ meV by keeping $\delta = 10$ meV and restricting to $\delta' < \delta$. In Fig. 15, we use $\delta' = 2.5$ meV.

7. Shift of peak positions in gap functions

Figure 15c. and d show the gap function Δ defined in Eqs. (3) and (A5). It reveals that the peak positions in $\Delta(k_{\text{AN}}, \omega)$ are different from those in $\text{Im}\Sigma^{\text{ano}}$, which is consistent with the hidden fermion theory¹⁶. In fact, the peak positions of $\text{Im}\Delta(k_{\text{AN}}, \omega)$ ($\sim \pm 80$ and ± 220 meV for Bi2212 and $\sim \pm 80$ and ± 210 meV for Bi2201) are nearly the same as the peak positions of $\text{Re}\Sigma^{\text{ano}}$, while the peak positions of $\text{Re}\Delta(k_{\text{AN}}, \omega)$ ($\sim \pm 50$ and ± 180 meV for Bi2212 and $\sim \pm 50$ and ± 160 meV for Bi2201) are nearly the same as the peak positions of $\text{Im}\Sigma^{\text{ano}}$. This is because the imaginary part of Q is dominant in the relevant frequency region ($\sim 100 - 200$ meV) as shown in Fig. 15. The shift of the peak positions indicates the strong renormalization effect in the normal quasiparticle contained in Q . In any case, in the contribution to the real order parameter of the superconductivity $\Delta_{\text{AN}} = \text{Re}\Delta(k_{\text{AN}}, \omega = 0)$ is expected to be contributed mostly from the two peaks in $\text{Im}\Delta(k_{\text{AN}}, \omega)$ through the Kramers-Kronig relation. The d -wave gap amplitude Δ_{AN} is 30 meV for the optimally doped Bi2212 while it is around

10 meV. However, at small energy (~ 60 meV), the gap amplitude is both around 40 meV, which is comparable to the peak energy of $\text{Im}\Sigma^{\text{ano}}$ and $\text{Im}\Sigma^{\text{nor}}$ suggesting the similar pseudogap energy for optimum and underdoped samples.

Appendix B: Robustness of machine learning

The present use of machine learning is categorized to a general class of regression analysis as addressed in the first paragraph of Appendix A. In the standard simple case of the regression task, training data set is simply given by the observed A at discrete number of x and we infer the functional form of $A(x)$. In the present case, it is more involved and the training data is the experimentally measured discrete and limited number of A and ω , and the regression task is to determine Σ as a continuous function of ω . In terms of the optimization with the machine learning, our task is to minimize the difference between the measured data A and that obtained from the inferred $\Sigma(\omega)$, which is a continuous function of ω . Therefore, our work is categorized to the machine learning application to a regression task, one of the most widely applied machine learning fields. Our regression scheme is illustrated in Figure 5

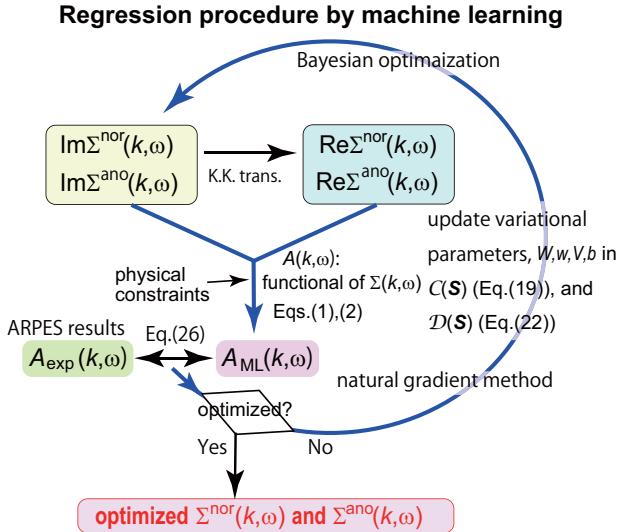


FIG. 5. Flow chart of machine-learning procedure in the present work The procedure to infer the normal self-energy, $\Sigma^{\text{nor}}(k, \omega)$ and anomalous self-energy $\Sigma^{\text{ano}}(k, \omega)$ using the experimental spectral function $A(k, \omega)$. K.K.trans. stands for Kramers-Kronig transformation. $A_{\text{ML}}(k, \omega)$ and $A_{\text{exp}}(k, \omega)$ are $A(k, \omega)$ determined from the machine learning and given by the ARPES measurement, respectively.

In the regression analysis, it is helpful to examine the reliability of the machine learning by using solvable cases as the benchmark, as in other type of the regression task

found in the problem of solving quantum many-body problems and classical statistical physics problems³². It is also important to test the stability of the procedure by adding noises. In this section we show the robustness against the noise and in Appendix F, we show several benchmark tests for solvable models.

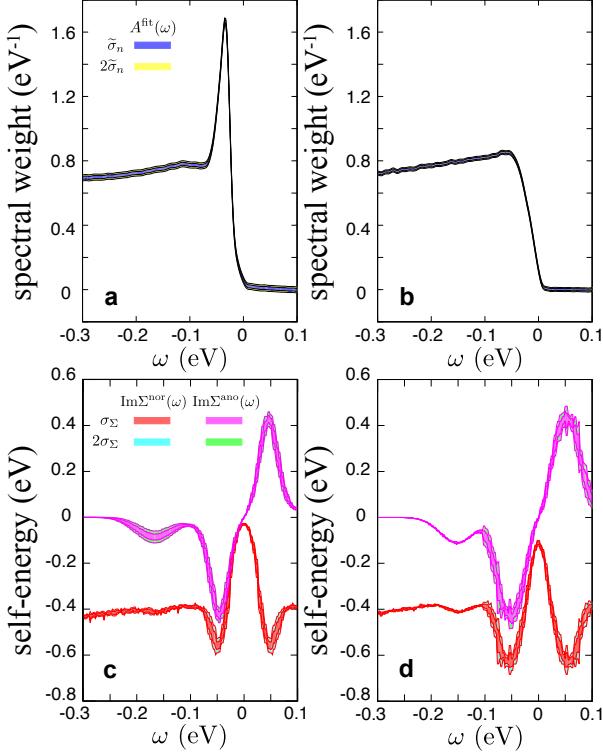


FIG. 6. Robustness of the machine-learning procedure
Examples of fictitious spectral weight with noise are shown for Bi2212 at 12 K (**a**) and for Bi2201 at 11 K (**b**), at the antinodal point. The black, grey, and light grey shaded area represent $A^{\text{fit}}(\omega) \pm \tilde{\sigma}_n$, $A^{\text{fit}}(\omega) \pm 2\tilde{\sigma}_n$, and $A^{\text{fit}}(\omega) \pm 3\tilde{\sigma}_n$. For the method of imposing noise, see Appendix A. Imaginary part of normal self-energy $\text{Im}\Sigma^{\text{nor}}(k_{\text{AN}}, \omega)$, $\text{Im}W(k_{\text{AN}}, \omega)$ and $\text{Im}\Sigma^{\text{tot}}(k_{\text{AN}}, \omega)$ deduced by the present machine learning from $A(k, \omega)$ including noise are shown for Bi2212 at 12 K (**a**) and for Bi2201 at 11 K (**b**). The belt width shows variance of the imaginary part of the self-energy, σ_{Σ} , arising from the variance of the fictitious noise, indicating the robustness of the present peak structure.

We examine stability of the present machine-learning scheme. By using A^{fit} and σ_n^2 , we can generate fictitious experimental data with larger amplitude of noise than the original data. Here, we use the fictitious data to examine the input data dependence of the present scheme. In the upper panel of Figure 6, we show examples of fictitious data with amplified noise $\tilde{\sigma}_n = 8\sigma_n$ for Bi2212 and $\tilde{\sigma}_n = 2\sigma_n$ for Bi2201. We perform the machine learning and extract the self-energy from the fictitious data. As shown in the lower panel of Fig. 6, the variance of $\text{Im}\Sigma^{\text{nor}}$, and $\text{Im}\Sigma^{\text{ano}}$ obtained from fictitious data is reasonably small

with the peak structure in the imaginary part of the self-energy, which indicates that our solution of the inverse problem is numerically stable.

The present machine-learning scheme is based on the imaginary parts of the self-energy within a finite frequency range $-\Lambda < \omega < \Lambda$, where $\Lambda \simeq 0.4$ eV, and the experimental data observed within $-0.4 \lesssim \omega \lesssim 0.2$. Therefore, in the genuine self-energy, there is a possible unknown contribution from the outside of the cutoff energy Λ . However, as explained below, such a contribution is a monotonic and bounded function of ω , and, thus, possible errors due to the lack of information can be estimated.

Due to the Kramers-Kronig relation, the real part of the self-energy can be affected by the cutoff energy Λ . Because the imaginary part of the normal self-energy is expected to extend over the cutoff energy, the real part of the normal self-energy has a monotonic and bounded contribution from the outside of the cutoff energy. On the other hand, because the anomalous self-energy is finite only within the cutoff energy scale, the real part of the anomalous self-energy can be affected by the cutoff only through the normal self-energy.

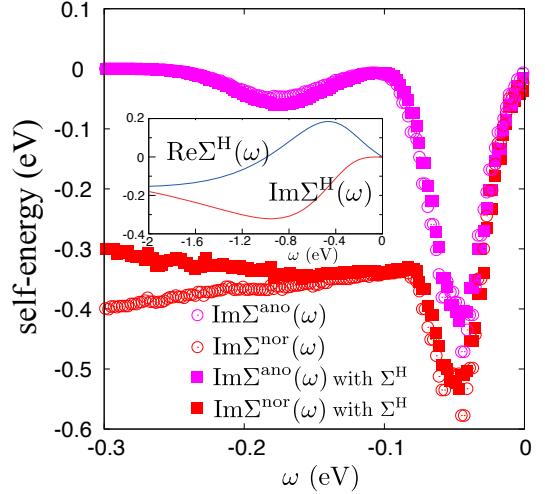


FIG. 7. Contribution from the high energy part to the self-energy structure. Artificial normal self-energy added by hand shown in the inset does not have appreciable effect for the self-energy near the fermi level as shown in the main panel. This artificially added normal self-energy is chosen not to violate the Kramers-Kronig relation.

In the main text, we ignored the contribution of the high-energy part of normal self-energy. To estimate the possible contribution from the outside of the cutoff energy, here, we assume a possible distribution of the imaginary part of the normal self-energy outside the cutoff: The imaginary part of the normal self-energy outside the cutoff is assumed to be confined within $\Omega' - W'/2 \lesssim \omega \lesssim \Omega' + W'/2$, where $|\Omega'| > \Lambda \gtrsim W'$ and the amplitude of the imaginary part is approximately constant within the

energy range. Then, contribution to the real part is given by

$$\Delta\text{Re}\Sigma^{\text{nor}}(\omega) \simeq \frac{\text{Im}\Sigma^{\text{nor}}(\Omega')}{\pi} \int_{\Omega'-W'/2}^{\Omega'+W'/2} \frac{d\omega'}{\omega - \omega'}, \quad (\text{B1})$$

which is monotonic for $-\Lambda < \omega < \Lambda$. When $|\Omega'| > \Lambda \gg W'$ is assumed, $\Delta\text{Re}\Sigma^{\text{nor}}(\omega = 0)$ and $\partial\Delta\text{Re}\Sigma^{\text{nor}}(\omega)/\partial\omega|_{\omega=0}$ are approximately estimated as $-\text{Im}\Sigma^{\text{nor}}(\Omega')W'/(\pi\Omega')$ and $\text{Im}\Sigma^{\text{nor}}(\Omega')W'/(\pi\Omega'^2)$, respectively. If we imagine formation of a lower Hubbard band, for instance, we may assume that $|\Omega'| \sim |\text{Im}\Sigma^{\text{nor}}(\Omega')| \sim \mathcal{O}(1)$ eV and $W' \ll \Lambda$. Then, the contribution from the outside of the cutoff is bounded as $|- \text{Im}\Sigma^{\text{nor}}(\Omega')W'/(\pi\Omega')| \ll \Lambda/\pi$ and $\text{Im}\Sigma^{\text{nor}}(\Omega')W'/(\pi\Omega'^2) \ll 1/(2\pi)$.

As we show in Fig. 7, the qualitative feature of the peak structure at $\omega < -0.4$ eV does not change even when we add an artificial high-energy part for the normal self-energy. In Fig. 7, we compare the self-energy inferred from the experimental ARPES data with the self-energy inferred with an additional high-energy part Σ^H .

Here, we added the following fixed high-energy part Σ^H in the optimization process,

$$\Sigma^H(\omega) = \frac{b}{(c-a)(b-c)} \left\{ \frac{\sqrt{b}}{\omega + i\sqrt{b}} - \frac{\sqrt{c}}{\omega + i\sqrt{c}} \right\} + \frac{a}{(a-b)(c-a)} \left\{ \frac{\sqrt{c}}{\omega + i\sqrt{c}} - \frac{\sqrt{a}}{\omega + i\sqrt{a}} \right\}, \quad (\text{B2})$$

where $a = 0.6$, $b = 0.5$, and $c = 0.3$, whose imaginary part becomes substantial for $\omega > -0.4$ eV as shown in the inset of Fig. 7.

Indeed, the artificial high-energy part does not affect the prominent peak structures in $\text{Im}\Sigma^{\text{nor}}$ and $\text{Im}\Sigma^{\text{ano}}$. When the energy range covered by the measured $A(k, \omega)$ becomes wider, the uncertainty becomes of course further reduced.

Appendix C: Comparison with Biased Self-Energy Form

Here, we discuss the possible origin of the discrepancy with the analysis by Bok *et al.*¹³. We first point out that the discrepancy is because of their neglect of the involved entanglement between Σ^{ano} and Σ^{nor} with a prominent peak in both of Σ^{ano} and Σ^{nor} revealed here. Their assumption of the fermi liquid for the normal component with momentum independence may not be justified. Since we are able to analyze with one EDC curve without assuming the fermi liquid behavior of the normal component, and successfully derive momentum dependent self-energy, the machine learning is more fit in solving the present problem, because the flexible fitting of the self-energy function is required.

In their analysis, (1) they have assumed momentum independent self-energy together with Eq.(S8) in their paper, which is justified only when the Fermi liquid theory applies in a wide energy range and as we clarified, it is not the case at least away from $\omega = 0$. (2) They have analyzed mainly only at $\theta = 20^\circ$, away from the antinodal point, which makes the identification of the prominent peak difficult. We infer that these may have effects on the contrast with ours.

Next we discuss the origin of discrepancy in the result by Li, *et al.*⁴². They assumed momentum independent self-energy in the MDC analysis as one sees in Eq.(3)

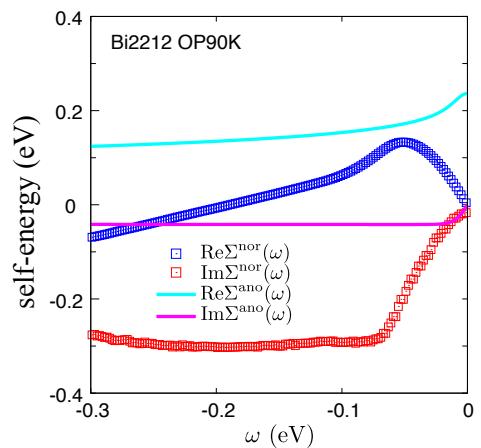


FIG. 8. Self-energies obtained by a biased anomalous self-energy. For details, see Appendix C.

of their supplementary note 2, whose basis is unclear. More crucially, they assumed the imaginary part of self-energy in the forms Eq.(5) or (6) for the normal part and (7) for the anomalous part, which is totally unjustified. Particularly, the form (7) does not allow the formation of peak and it does not allow the cancellation with the structure in the normal contribution in the spectral function as we discovered. It is crucially important to allow the flexibility of the self-energy form and the machine learning is one of the best way to incorporate it while the attempt by Li, *et al.* failed it. We have attempted to fit the self-energy in the form of (7) for the anomalous part and found the resultant optimized χ^2 is 1.71×10^{-4} , which is higher than the present result 1.15×10^{-4} . The

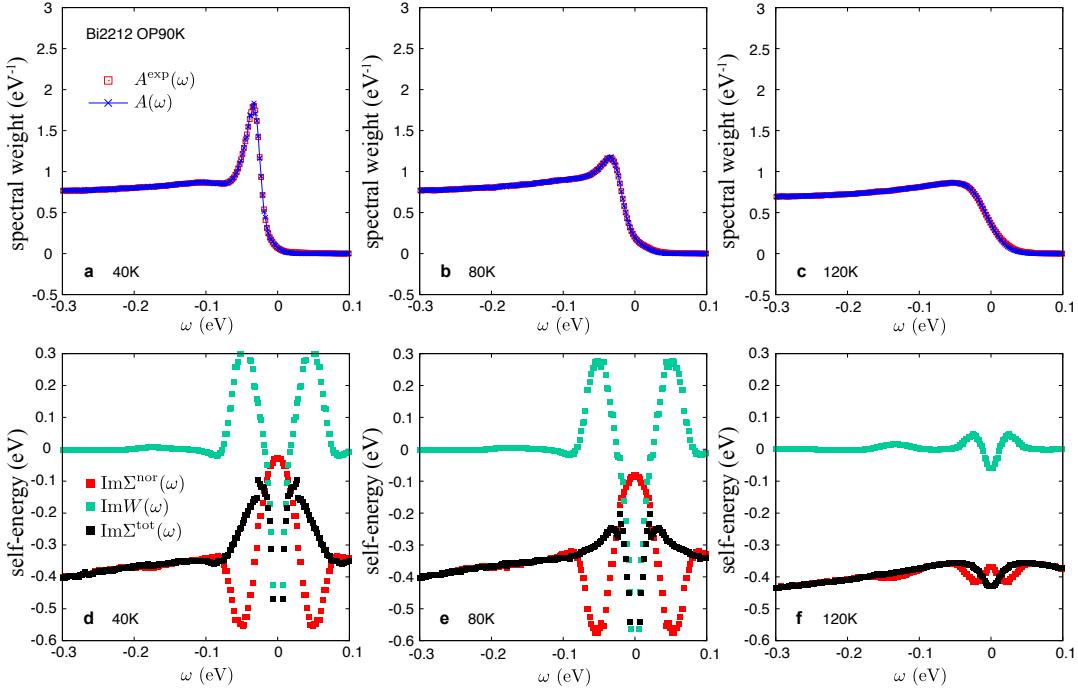


FIG. 9. Temperature dependence of self-energies.

Self-energies are obtained from the machine learning using the ARPES EDC curves plotted in the upper panel and taken from the experimental data of Bi2212 at optimum doping at 40K, 80K and 120K, which are supplied by Kondo *et al.*¹⁰. Although the peak cancellation between W and Σ^{nor} still exists at 40K and 80K, above T_c the anomalous self-energy peak contribution to W essentially vanishes, as is expected.

resultant self-energy does not show any appreciable peak as it should be in contrast to the present result as one sees in Fig. 8. Note that due to the unbiased choice of the normal self-energy here, the self-energies in Fig. 8 are relatively better fit than the more biased ones in Li, *et al*⁴². The spectral function obtained from the additional bias in the normal part, Eq. (5) or (6) in Ref. 42 must give higher χ^2 .

Appendix D: Machine Learning Results above T_c

We have shown the machine learning results in the main text for the superconducting phase well below T_c to show the remarkable structure of the pronounced anomalous self-energy peak with its dominant contribution to the superconductivity. However, the question how the cancellation of the normal and anomalous self-energy evolves with raising temperatures provides us with further insight on its role in the superconductivity. Fig. 9 shows machine learning results of the normal and anomalous self-energies together with W for the ARPES measurement obtained by the machine learning of Bi2212 at optimum doping at 40, 80 and 120K¹⁰. The anomalous self-energy peak vanishes above T_c as it should be, which further confirms the validity of the present machine learn-

ing scheme.

Appendix E: Momentum Dependence

We have shown the machine learning results in the main text at the antinodal point, because the remarkable structure of the pronounced anomalous self-energy peak, coexisting with the normal self-energy peak is most clearly identified with its dominant contribution to the superconductivity. However, the momentum dependence of the peak structure provides us with useful insight. We here show the momentum dependence of self-energy structure. Fig. 10 shows the imaginary part of the normal and anomalous self-energies together with W for the ARPES measurement angle 1.0° , 11.1° , 21.2° , 31.3° , and 41.9° obtained by the machine learning result of Bi2201 at optimum doping at 11.3 K¹⁰. Note that 0° is the antinodal and 45° is the nodal points. Although the peaks in the normal and anomalous self-energy become less significant as is expected, the cancellation of $\text{Im}\Sigma^{\text{nor}}$ and W always holds and the prominent peak is missing in $\text{Im}\Sigma^{\text{tot}}$ similarly to the case at the antinodal point. This result further corroborates the universal mechanism of the peak cancellation and the dominant contribution to the superconductivity.

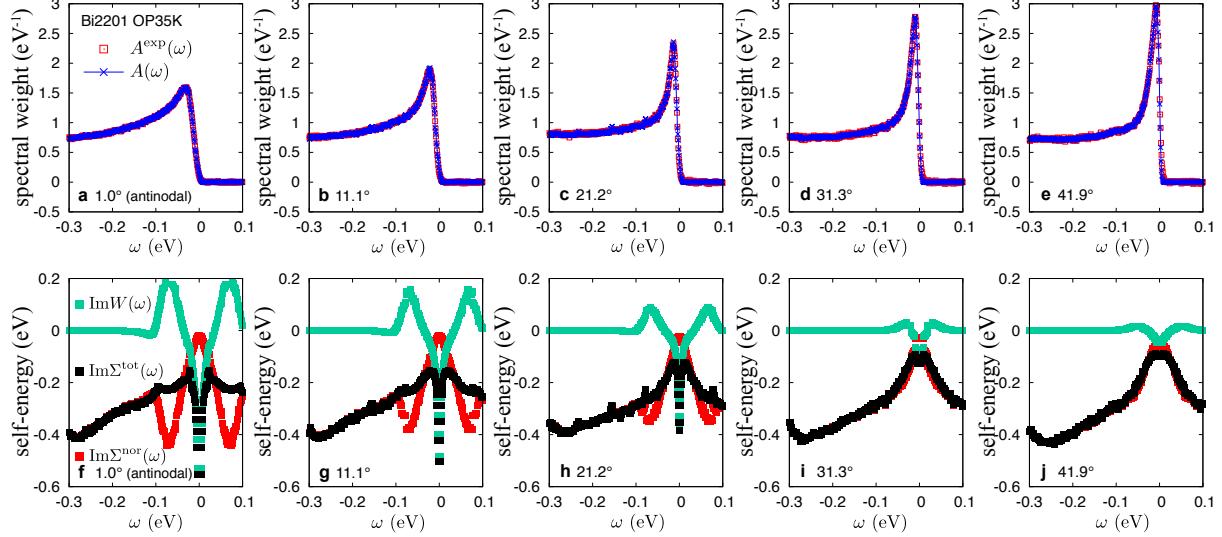


FIG. 10. Momentum (angle) dependence of self-energies.

Self-energies are obtained from the machine learning using the ARPES EDC curves plotted in the upper panel and taken from the experimental data of Bi2201 at optimum doping at 11.3 K at the angle 1.0° (**a.** and **f.**), 11.1° (**b.** and **g.**), 21.2° (**c.** and **h.**), 31.3° (**d.** and **i.**), and 41.9° (**e.** and **j.**) supplied by Kondo *et al.*¹⁰. Although the quasiparticle peak becomes sharper when the nodal point is approached, prominent peaks are found at all angles in imaginary parts of the normal and anomalous self-energies around ± 0.07 eV, which are missing in W at all the angles, though the peaks become less pronounced and are almost missing at 41.9° (nearly nodal point).

Appendix F: Accuracy, Stability and Robustness of the Present Method Clarified by Solvable Benchmarks

In this section, we employ exactly solvable models as benchmarks. We first examine whether our machine learning correctly reproduces the exact self-energies (with prominent peak structures), if the exact solution indeed shows the cancellation of the normal and anomalous self-energy contributions in the total self-energy and the spectral function $A(\omega)$ shows only a weak peak-dip-hump structure. To our knowledge, exact solution, which shows such a cancellation is not found except for the case of the two-component fermion model. Then, as a benchmark, we inferred the self-energy of a superconducting two-component fermion model defined by the following Lagrangian,

$$L(\omega) = \sum_{k,\sigma} [(\omega + i\delta - \epsilon_c(k) - \Sigma^{(0)}(\omega))c_{k,\sigma}^\dagger c_{k,\sigma} - \epsilon_f f_{k,\sigma}^\dagger f_{k,\sigma} - V_1(c_{k,\sigma}^\dagger f_{k,\sigma} + \text{H.c.}) - D_1(f_{k,\sigma}^\dagger f_{-k,-\sigma}^\dagger + \text{H.c.})], \quad (\text{F1})$$

which is essentially the same form as that introduced in Ref.16. In the following discussion, we assume that the noninteracting density of states determined from $\epsilon_c(k)$ is a constant N_F and focus on a specific momentum k at the Fermi momentum just for simplicity. Because of the momentum independence, this consideration at a specified momentum does not cause loss of generality. Here,

we add $\Sigma^{(0)}(\omega)$ at the above momentum defined by

$$\Sigma^{(0)}(\omega) = \frac{d}{b-a} \left\{ \frac{\sqrt{b}}{\omega + i\sqrt{b}} - \frac{\sqrt{a}}{\omega + i\sqrt{a}} \right\} \quad (\text{F2})$$

in addition to $\epsilon_c(k)$ to mimic the additional normal Fermi-liquid-like component seen in the experimental result arising from interaction effect for the part not represented by the coupling to the f fermion, where a , b and d are constants. The self-energy in the exact solution of this two-fermion model is given as,

$$\Sigma_{2f}^{\text{nor}}(\omega) = \frac{V_1^2(\omega + i\delta + \epsilon_f)}{(\omega + i\delta)^2 - (\epsilon_f^2 + D_1^2)} + \epsilon_c + \Sigma^{(0)}(\omega), \quad (\text{F3})$$

$$\Sigma_{2f}^{\text{ano}}(\omega) = -\frac{V_1^2 D_1}{(\omega + i\delta)^2 - (\epsilon_f^2 + D_1^2)}, \quad (\text{F4})$$

For simplicity, we have dropped the momentum dependence in the solutions (F3) and (F4). In our calculation, we set $a = 0.008$ eV², $b = 0.2$ eV², $d = 0.08$ eV³, $V_1 = 0.075$ eV, $D_1 = 0.0375$ eV, and $\epsilon_f = \epsilon_c = 0$. The present choice of the parameters is enough to generate the spectral function observed at the Fermi momentum we focus, and the dependence on the doping and dimension of the system *etc.* are implicitly contained in N_F . By using the exact solution for the spectral function $A_{2f}(\omega)$, we add small but finite noise, where σ^2 of the noise is set to be 6×10^{-4} to mimic the experimental noise and perform the machine learning using this noisy $A_{2f}(\omega)$.

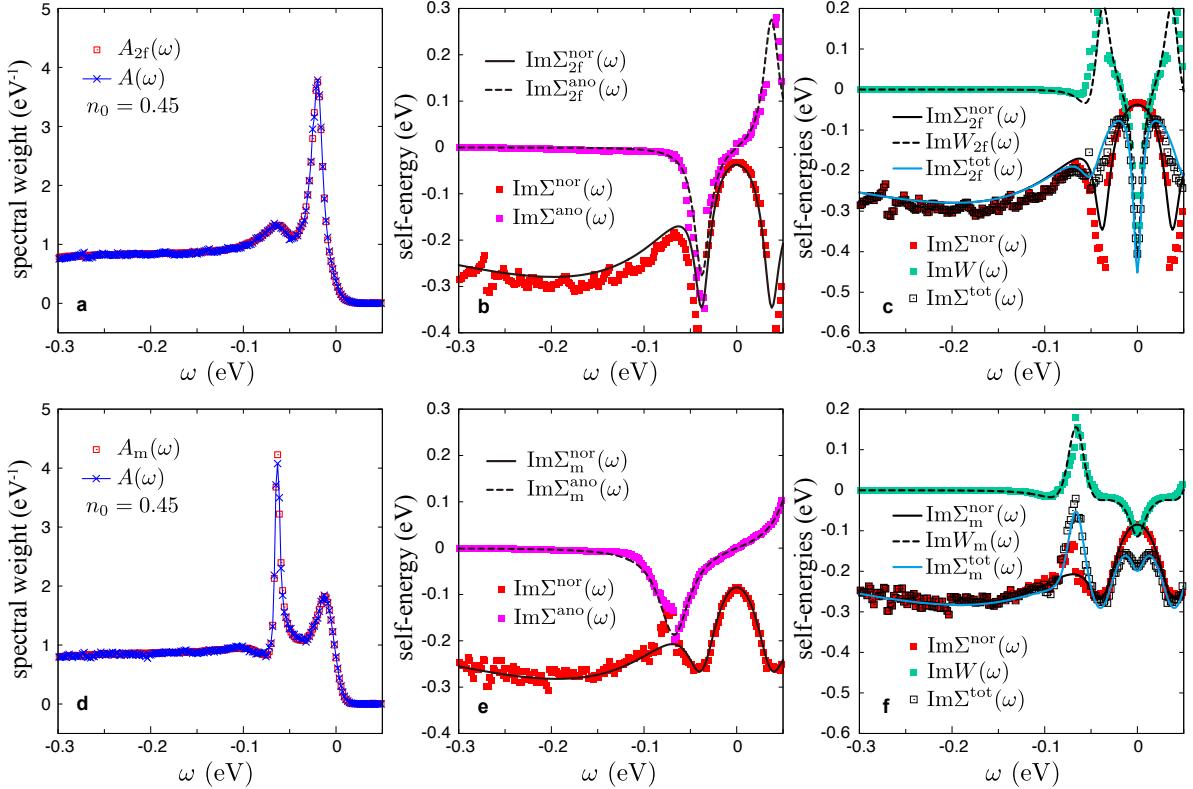


FIG. 11. **Spectrum and self-energies of two-component fermion model and its modified version.** **a.** The exact spectral weight of the two-component fermion model, $A_{2f}(\omega)$, (red open squares) is compared with $A(\omega)$ obtained from the machine learning using $A_{2f}(\omega)$ (blue curve crosses) within $-0.55 \text{ eV} < \omega < 0.05 \text{ eV}$. **b.** and **c.** The imaginary part of the self-energies of the two-component fermion model, Σ_{2f}^{nor} , Σ_{2f}^{ano} , and W_{2f} (curves), are compared with the self-energies obtained from the machine learning (symbols). **d.** The exact spectral weight of the modified two-component fermion model, $A_m(\omega)$, (open squares) is compared with $A(\omega)$ obtained from the machine learning using $A_m(\omega)$ within $-0.55 \text{ eV} < \omega < 0.05 \text{ eV}$. **e.** and **f.** The imaginary part of the self-energies of the modified two-component fermion model, Σ_m^{nor} , Σ_m^{ano} , and W_m (curves), are compared with the self-energies obtained from the machine learning (symbols).

Although it is irrelevant to the inferred self-energies, the Fermi-Dirac distribution with $T = 40 \text{ K}$ is introduced in the spectral weight used in the machine learning just by following the scheme used in the main text. In Figs. 11a-c, the spectral function of the two-component fermion model $A_{2f}(\omega)$ and the self-energies are shown for the exact solutions (solid and dashed curves) and the machine learning results (symbols). In the self-energy inference, we choose $n_0 = 0.45$ without fine tuning and use the spectrum within $-0.55 \text{ eV} < \omega < 0.05 \text{ eV}$. The peak position in $\text{Im}\Sigma^{\text{nor}}$ and $\text{Im}\Sigma^{\text{ano}}$ and the Fermi liquid-like normal contribution in the exact solution (shown with the index $2f$ such as $A_{2f}(\omega)$ illustrated by solid and dashed curves) are well reproduced by our machine learning results (symbols). The peak cancellation in Σ_{2f}^{tot} in the exact solution is also well reproduced.

In Fig. 11d-f, we show an artificial case, where the pole of Σ^{ano} is shifted 0.03 eV from the solution (F4), where the pole cancellation in the total self-energy does not occur any more and the spectral weight shows weird two peaks. The concrete representation of the self-energies

of the modified two-component model is given by the following form,

$$\Sigma_m^{\text{nor}}(\omega) = \frac{V_1^2(\omega + ir\delta + \epsilon_f)}{(\omega + ir\delta)^2 - (\epsilon_f^2 + D_1^2)} + \epsilon_c + \Sigma^{(0)}(\omega), \quad (F5)$$

$$\Sigma_m^{\text{ano}}(\omega) = -\frac{V_1^2(D_1 + \Delta D_1)}{(\omega + ir\delta)^2 - (\epsilon_f^2 + (D_1 + \Delta D_1)^2)}, \quad (F6)$$

where the pole of Σ^{ano} is shifted from that of Σ^{ano} because of $\Delta D_1 = 0.03 \text{ eV}$ and a factor $r = 2$ is introduced to avoid singular spectrum. The spectrum and self-energies of the modified two-component model are denoted with the index m as $A_m(\omega)$. Even in this case, the machine learning results well reproduce all the line shapes. This indicates that our machine learning flexibly and accurately reproduces the exact solution irrespective of the presence or absence of the peak cancellation.

The present scheme also reproduces the self-energies of boson-mediated superconductors. As a typical model,

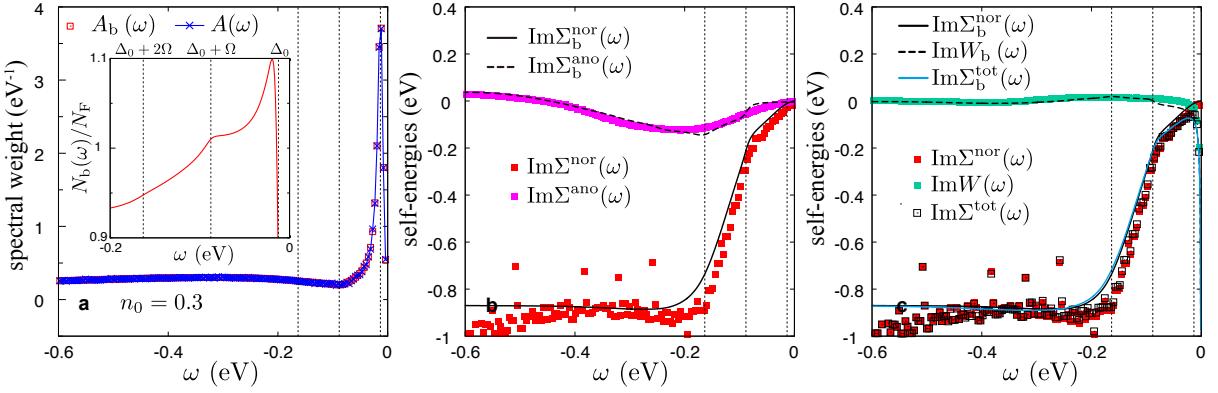


FIG. 12. Spectrum and self-energies of phonon-mediated superconducting model. **a.** The exact spectral weight of the phonon-mediated superconductor, $A_b(\omega)$, (red open squares) is compared with $A(\omega)$ (blue curve with crosses) obtained from the machine learning using $A_b(\omega)$ within $-1 \text{ eV} < \omega < 0.0 \text{ eV}$. The inset shows the electronic density of states in the superconducting state normalized by the normal state density of states N_F . The vertical dotted lines show $\omega = -\Delta_0$, $-\Delta_0 - \Omega$, and $-\Delta_0 - 2\Omega$, where anomalies appear reflecting strong electron-phonon couplings and formation of the superconducting gap Δ_0 . While the spectral weight at the Fermi momentum $A_b(\omega)$ shows a dip at $\omega = -\Delta_0 - \Omega$, the density of state shows a shoulder due to the so-called kink in the renormalized dispersion that appears when a finite energy shift ϵ is introduced. **b.** and **c.** The imaginary part of the self-energies of the phonon-mediated superconductor, $\Sigma_b^{\text{nor}}(\omega)$, $\Sigma_b^{\text{ano}}(\omega)$, $\Sigma_b^{\text{tot}}(\omega)$ and $W_b(\omega)$ (curves), are compared with the self-energies obtained from the machine learning (symbols). Note that the curves for Σ_b^{tot} and Σ_b^{nor} are nearly overlapped in **c.**

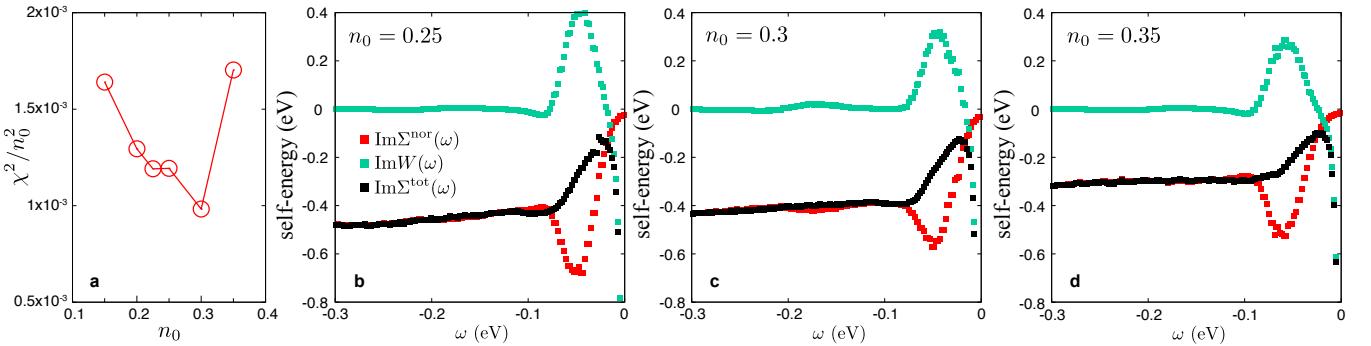


FIG. 13. Determination of n_0 . In the main text, we have employed $n_0 = 0.3$. To assess the validity of this choice, we show n_0 dependence of χ^2 defined in Eq.(A30).

the superconducting state described by the following Eliashberg equations is examined.

By following the standard strong coupling theory for boson-mediated superconductors^{5,43,44}, the superconducting gap $\Delta(\omega)$ and particle-hole symmetric component of the normal self-energy $\omega(1 - Z(\omega))$ at zero temperature are given by the Eliashberg equations,

$$\Delta(\omega) = \frac{N_F}{Z(\omega)} \int_{\Delta_0}^{\omega_c} d\omega' \text{Re} \left\{ \frac{\Delta(\omega')}{[\omega'^2 - \Delta(\omega')^2 + i\eta]^{1/2}} \right\} K_+(\omega, \omega'), \quad (\text{F7})$$

$$[1 - Z(\omega)] \omega = N_F \int_{\Delta_0}^{\infty} d\omega' \text{Re} \left\{ \frac{\omega'}{[\omega'^2 - \Delta(\omega')^2 + i\eta]^{1/2}} \right\} K_-(\omega, \omega') + \Sigma^{(0)}(\omega), \quad (\text{F8})$$

where $\Delta_0 = \text{Re}\Delta(\Delta_0)$, ω_c is the cutoff frequency, and η is a positive broadening factor.

Again, we assume that the noninteracting density of states is given by momentum and energy independent

constant N_F for simplicity and the superconducting symmetry is momentum independent s -wave. In the following, we use the name “phonon” for the boson, although any kind of boson can be the candidate of this treatment. Here, we assume that the kernel functions K_{\pm} originate from the Einstein phonon as

$$K_{\pm}(\omega, \omega') = g^2 \left[\frac{1}{\omega' + \omega + \Omega - i\eta} \pm \frac{1}{\omega' - \omega + \Omega - i\eta} \right], \quad (\text{F9})$$

where Ω is the Einstein phonon frequency and g is the electron-phonon coupling constant. With the assumption that the density of states is constant around the Fermi level, the self-energies obtained by the Eliashberg equations are independent of the electron density and dimension of the system. Then, the normal component of the Green’s function is defined as

$$G_b^{\text{nor}}(\epsilon, \omega) = \frac{Z(\omega)\omega + \epsilon}{\{Z(\omega)\omega\}^2 - \epsilon^2 - \phi(\omega)^2 + i\eta}, \quad (\text{F10})$$

where $\phi(\omega) = Z(\omega)\Delta(\omega)$ and ϵ is the energy measured from the Fermi energy. The spectral function is defined as $A_b(\omega) = -(1/\pi)f_{\text{FD}}(\omega)\text{Im}G_b^{\text{nor}}(\epsilon = 0, \omega)$, where $f_{\text{FD}}(\omega)$ is the Fermi-Dirac distribution, and the superconducting density of states is given by

$$N_b(\omega) = N_F \int d\epsilon |-(1/\pi)\text{Im}G_b^{\text{nor}}(\epsilon, \omega)|. \quad (\text{F11})$$

The notation of the self-energies used in the literature on the phonon-mediated superconductors^{43,44}, $\omega[1 - Z(\omega)]$ and $\phi(\omega)$, is different from $\Sigma^{\text{nor}}(\omega)$ and $\Sigma^{\text{ano}}(\omega)$ in the present paper. The normal and anomalous components of the self-energies are obtained as

$$\Sigma_b^{\text{nor}}(\omega) = \omega - Z(\omega)\omega, \quad (\text{F12})$$

$$\Sigma_b^{\text{ano}}(\omega) = \phi(\omega) = Z(\omega)\Delta(\omega). \quad (\text{F13})$$

In Fig. 12, an example of the self-energy inference for the phonon-mediated superconductors is shown. Here, we set the coupling constant $g^2 N_F = 0.275$ eV, the Einstein phonon frequency $\Omega = 0.075$ eV, the cutoff frequency $\omega_c = 4$ eV, and the broadening factor $\eta = 0.0075$ eV². The parameters in $\Sigma^{(0)}(\omega)$ are chosen as $a = 0.008$ eV², $b = 0.016$ eV², and $d = 0.005$ eV³. For normalization of the spectral function, we choose $n_0 = 0.3$ without tuning. Although its effect is negligibly small in the inferred self-energies, the Fermi-Dirac distribution with $T = 40$ K is introduced in the spectral weight used in the machine learning just by following the scheme used in the main text. The machine learning results capture essential features of the exact normal and anomalous self-energies, where the anomalous self-energy has a dip around $-(\Delta_0 + \Omega)$ and $-(\Delta_0 + 2\Omega)$ and the normal self-energy shows a sharp drop. The dip in the anomalous self-energy, which arises from the electron-phonon coupling and gives the superconducting gap through the Kramers-Kronig relation, is responsible for the s -wave

superconductivity. These features are characteristic of the strong coupling BCS (phonon-mediated) superconductors and the anomaly of $N_b(\omega)$ at $-(\Delta_0 + \Omega)$ shown in the inset of a. is regarded as the evidence of the phonon mechanism. The machine learning well reproducing the exact results of the dip in $\text{Im}\Sigma^{\text{ano}}$ and the sharp drop in $\text{Im}\Sigma^{\text{nor}}$ indicates the reliability of the present method.

Appendix G: Gaussian Distribution Represented by Boltzmann Machine

When we choose the parameters as

$$V_{\ell m}^{\lambda} = -\frac{1}{2s_{\lambda}^2} \left(\frac{\Lambda}{2^L} \right)^2 2^{\ell+m}, \quad (\text{G1})$$

$$b_{\ell}^{\lambda} = \frac{1}{s_{\lambda}^2} (\Lambda/2 + x_{\lambda}) \frac{\Lambda}{2^L} 2^{\ell}, \quad (\text{G2})$$

$$w_{\lambda} = \frac{w_{0\lambda}}{\sqrt{2\pi s_{\lambda}^2}} e^{-\frac{1}{2s_{\lambda}^2}(x_{\lambda} + \Lambda/2)^2}, \quad (\text{G3})$$

the Boltzmann machine easily represents the Gaussian distribution with the center x_{λ} , variance s_{λ}^2 , and weight $w_{0\lambda}$, which is a localized sparse distribution. Superposition of the Gaussian distribution can easily be expressed by Eq.(A20) by taking $1 < M < 4$.

Since it is sufficient to take the number of variables S_{ℓ} at most 9 to fit the experiment data containing the resolution limitation, we can explicitly take the trace summation over S_{ℓ} for all ℓ (with 2^L terms) at each iteration step. Therefore the drawback of the form Eq.(A20) beyond the RBM (namely, the complexity arising from containing the term proportional to $V_{\ell m}^{\lambda}$) is not a serious problem.

Appendix H: Optimization of n_0

We employed $n_0 = 0.3$ in the present analyses. This choice is justified by the least square fit. We have examined the optimum choice by the least square fit of χ^2 by taking several choices of n_0 . Here, we note that χ^2 is trivially scaled by the square of the amplitude of $A^{\text{exp}}(\omega)$, and thus is scaled by the square of n_0 . Therefore, we need to optimize χ^2/n_0^2 . Fig. 13a shows that χ^2 normalized by n_0^2 for optimally doped Bi2212 at the antinodal momentum has indeed minimum at $n_0 = 0.3$, which indicates that the machine learning suggests that this choice is the optimized value of n_0 . The obtained self-energies do not sensitively depend on the choice of n_0 as one sees in Figs. 13b-d and we see no qualitative change in the feature of the pronounced peaks in $\text{Im}\Sigma^{\text{ano}}$ and $\text{Im}\Sigma^{\text{nor}}$ at the same energy together with their cancellation in $\text{Im}\Sigma^{\text{tot}}$.

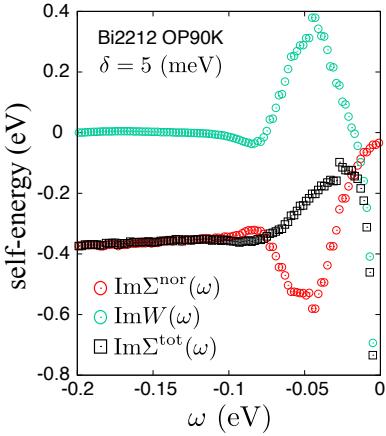


FIG. 14. **Resolution δ dependence of self-energies.** Self-energies are obtained from the machine learning using the ARPES EDC curves with $\delta=5$ meV. The ARPES EDC is taken from the experimental data of Bi2212 at optimum doping at 11K, which are supplied by Kondo *et al.*⁹. The peak position and their cancellation between W and Σ^{nor} remain essentially the same even for δ smaller than the experimental resolution.

Appendix I: Effects of Resolution δ

In the present study, the small imaginary part $i\delta$ utilized in the Green's functions is chosen to be equal to the experimental resolution. When substantially larger resolution δ is taken, the detailed spectra are trivially not reproducible. On the other hand, when smaller resolution δ is taken, the spectra may be easily fitted. Here, we examine how the smaller δ affects the inferred self-energy. As a typical example, we take $\delta=5$ meV, which is a half of δ used in the main article, and confirmed that the smaller δ does not change the qualitative structure of the self-energy. As shown in Fig. 14, the peak structures in $\text{Im}\Sigma^{\text{nor}}$ and $\text{Im}W$, and the cancellation between them are reproduced.

Appendix J: Pair Breaking in Underdoped Sample

The renormalization factor (quasiparticle residue) estimated from the expression

$$z(k) = Q(k, \omega)|_{\omega \rightarrow 0} \quad (\text{J1})$$

theoretically equivalent to z_{qp} defined in Eq.(A7) is the weight of the quasiparticle, which can be substantially reduced from the noninteracting value $z(k) = 1$ due to the interaction effects. The renormalization factor estimated from the fitting of Eq.(J1) is $z_{\text{qp}} \sim 0.1$ for Bi2212 and $z_{\text{qp}} \sim 0.03$ for Bi2201 (see Fig. 15) supporting the non-Fermi liquid behavior especially in the underdoped case.

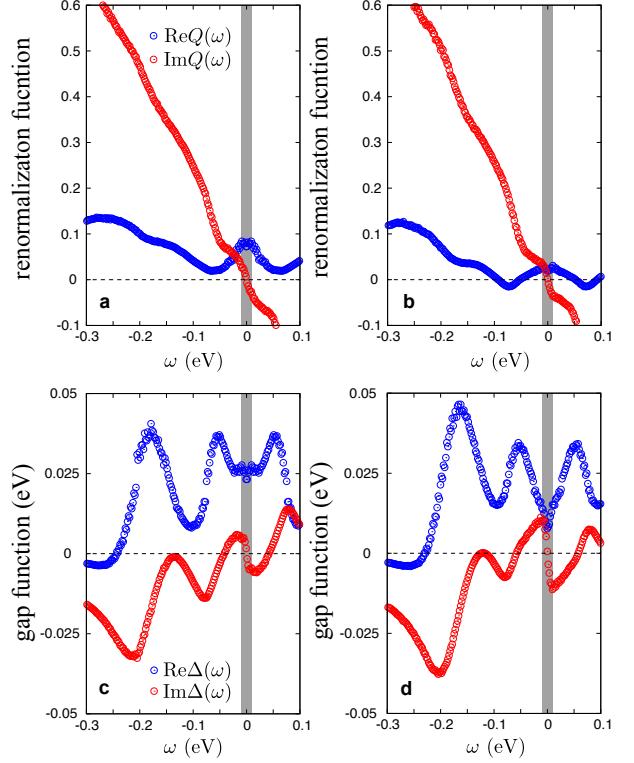


FIG. 15. **Renormalization function $Q(k_{\text{AN}}, \omega)$ and Gap function $\Delta(k_{\text{AN}}, \omega)$ obtained by machine learning.** $Q(k_{\text{AN}}, \omega)$ (a. and b.) and $\Delta(k_{\text{AN}}, \omega)$ (c. and d.) are plotted for optimally doped Bi2212 (a. and c.) and underdoped Bi2201 (b. and d.) for the experimental data shown in Fig. 1. The width of shaded area shows the experimental resolution (see Method). The positive part in $\text{Im}\Delta(k_{\text{AN}}, \omega)$ near $\omega = 0$ indicates the pair breaking contributing to destroy the superconductivity.

As shown in Figs. 15c and d, the non-Fermi-liquid-like $\text{Im}\Sigma^{\text{nor}}(\omega)$ affects the gap function $\Delta(\omega)$ (in Eq.(3)) through $Q(\omega)$ (see Eq.(A5)). In general, negative $\text{Im}\Delta(\omega)$ for $\omega < 0$ enhances $\text{Re}\Delta(\omega = 0)$ through the Kramers-Kronig relation and is indeed negative in most of ω in Fig. 15c, d. However, $\text{Im}\Delta(\omega)$ is positive at $|\omega| < 0.04$ eV ($|\omega| < 0.06$ eV) for Bi2212 (Bi2201). Because $\text{Im}\Sigma^{\text{ano}}$ is found to be always negative for $\omega < 0$, it is ascribed to the pair breaking effect of Q , arising from poles of Σ^{nor} inside the superconducting gap as already pointed out¹⁶. The pair breaking is much more prominent for underdoped sample, Bi2201.

Although a similar conclusion for the underdoped Bi2201 suggests a universal nature, the prominent non-Fermi liquid behavior and the pair breaking could be accounted for by an alternative at $k = k_{\text{AN}}$, namely the pole of Σ^{nor} shifts to the energy $\omega \sim 0$ and destroys Σ^{ano} accompanied by an insulating gap. Although such a solution gives worse χ^2 in our analysis, a momentum selective insulating behavior at the antinodal point deserves to be explored further together with the full momentum and

temperature dependences.

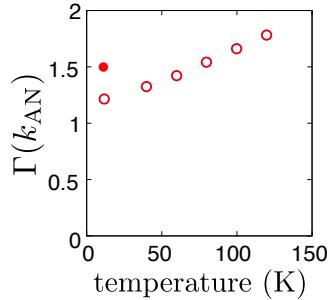


FIG. 16. Single-particle relaxation time derived from machine learning. Temperature dependence of $z_{\text{qp}}c_1$ of the underdoped Bi2201 with $T_c = 23$ K at k_{AN} . At the lowest temperature, the results of two different samples are shown: The closed circle denotes $z_{\text{qp}}c_1$ inferred from the underdoped Bi2201 reported in Ref. 9 and the open circles denote $z_{\text{qp}}c_1$ inferred from the underdoped Bi2201 reported in Ref. 10.

Appendix K: Local and temperature insensitive scattering rate $z_{\text{qp}}(k)c_1(k)$

Angle (momentum) dependence of $z_{\text{qp}}(k_F)c_1(k_F)$ plotted for Bi2201 in Fig. 4 and resultant $\tau(k)$ shows that it is only weakly dependent around the unity on the angle and doping concentration. (See also Fig. 17 for the plots for each $z_{\text{qp}}(k_F)$ and $c_1(k_F)$.) Even for the optimal Bi2212 at the antinodal point, despite the large difference in T_c , the value of $z_{\text{qp}}(k)c_1(k)$ is similar (~ 1.4). (Note that the value $z_{\text{qp}}(k_F)c_1(k_F)$ is somewhat large (~ 1.5) at the nodal point for the underdoped Bi2201 sample, consistently with the increasing slope of the T -linear resistivity in the underdoped region²⁵. This could be related to the effect of competing insulating behavior.)

With raising temperatures, experimentally observed

T -linear resistivity has to lead to the temperature insensitive $z_{\text{qp}}(k)c_1(k)$ at least near the node, because the transport is governed by the nodal region. Temperature dependence of $z_{\text{qp}}(k)c_1(k)$ in the normal state at k_{AN} , for instance for the underdoped Bi2201 is also weak with a large constant offset shown in Fig. 16, which implies that the T -linear dependence ($\propto z_{\text{qp}}(k)c_1(k)T$) is preserved irrespective of the momentum. It supports the local nature of dissipation saturated against temperature below and above T_c and intrinsically quantum mechanical.

Appendix L: Momentum and doping dependences of $F(k)$, $z_{\text{qp}}(k)$, $c_1(k)$ and $\Delta_{\text{qp}}(k)$

In Fig. 17, doping concentration dependences of the superconducting carrier density $F(k)$, the mass renormalization factor $z_{\text{qp}}(k)$, $c_1(k)$ defined as the ω -linear component of $\text{Im}\Sigma^{\text{nor}}(k, \omega)$, and $\Delta_0(k)$ at the Fermi momentum k_F are plotted. Here, $c_1(k)$ is defined by

$$c_1(k) = \frac{\partial \text{Im}\Sigma^{\text{nor}}(k, \omega)}{\partial \omega} |_{\omega \sim 0} \quad (\text{L1})$$

obtained from the linear fitting of $\text{Im}\Sigma^{\text{nor}}(k, \omega)$ in the range of $15\text{meV} < \omega < 40\text{meV}$.

Discrepancy between the doping dependence of T_c and quasiparticle gap amplitude, established in the literature^{29,45}, is further examined by the present self-energy learning. In Table I, the doping dependences of the density of Cooper pairs $F(k)$, the gap amplitude estimated from the peak position in EDC, the superconducting gap Δ_0 obtained by the Boltzmann machine learning, the quasiparticle renormalization factor z_{qp} , $c_1(k)$, \bar{Q} , P , ω_{PEAK} and Ω_0 obtained from ARPES data of underdoped (UD), optimally doped (OP), and overdoped (OD) Bi2201 samples¹⁰ are summarized. While the order parameter F and superconducting gap Δ_0 shows dome-like doping dependence as T_c does, the gap amplitude estimated from the peak position in EDC monotonically decreases upon increasing doping^{29,45}. On the other hand, $z_{\text{qp}}(k_{\text{AN}})$ monotonically increases.

* yamaaji@ap.t.u-tokyo.ac.jp

¹ A. Damascelli, Z. Hussain, Z. X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003).

² M. Norman, H. Ding, H. Fretwell, M. Randeria, J. Campuzano, *Physical Review B* **60**, 7585 (1999).

³ S. Aizaki, et al., *Phys. Rev. Lett.* **109**, 056401 (2012).

⁴ J. E. Hoffman, et al., *Science* **297**, 1148 (2002).

⁵ W. L. McMillan, J. M. Rowell, *Phys. Rev. Lett.* **14**, 108 (1965).

⁶ J. M. Rowell, W. L. McMillan, W. L. Feldmann, *Phys. Rev. B* **3**, 4065 (1971).

⁷ D. H. Ackley, G. E. Hinton, T. J. Sejnowski, *Cognitive science* **9**, 147 (1985).

⁸ P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory, *Tech. rep.*, COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE (1986).

⁹ T. Kondo, R. Khasanov, T. Takeuchi, J. Schmalian, A. Kaminski, *Nature* **457**, 296 (2009).

¹⁰ T. Kondo, et al., *Nature Physics* **7**, 21 (2011).

¹¹ L. Kadanoff, G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).

¹² D. J. Scalapino, J. R. Schrieffer, J. W. Wilkins, *Phys. Rev.* **148**, 263 (1966).

¹³ J. M. Bok, et al., *Science Advances* **2**, e1501329 (2016).

¹⁴ M. R. Norman, M. Randeria, H. Ding, J. C. Campuzano, *Phys. Rev. B* **57**, R11093 (1998).

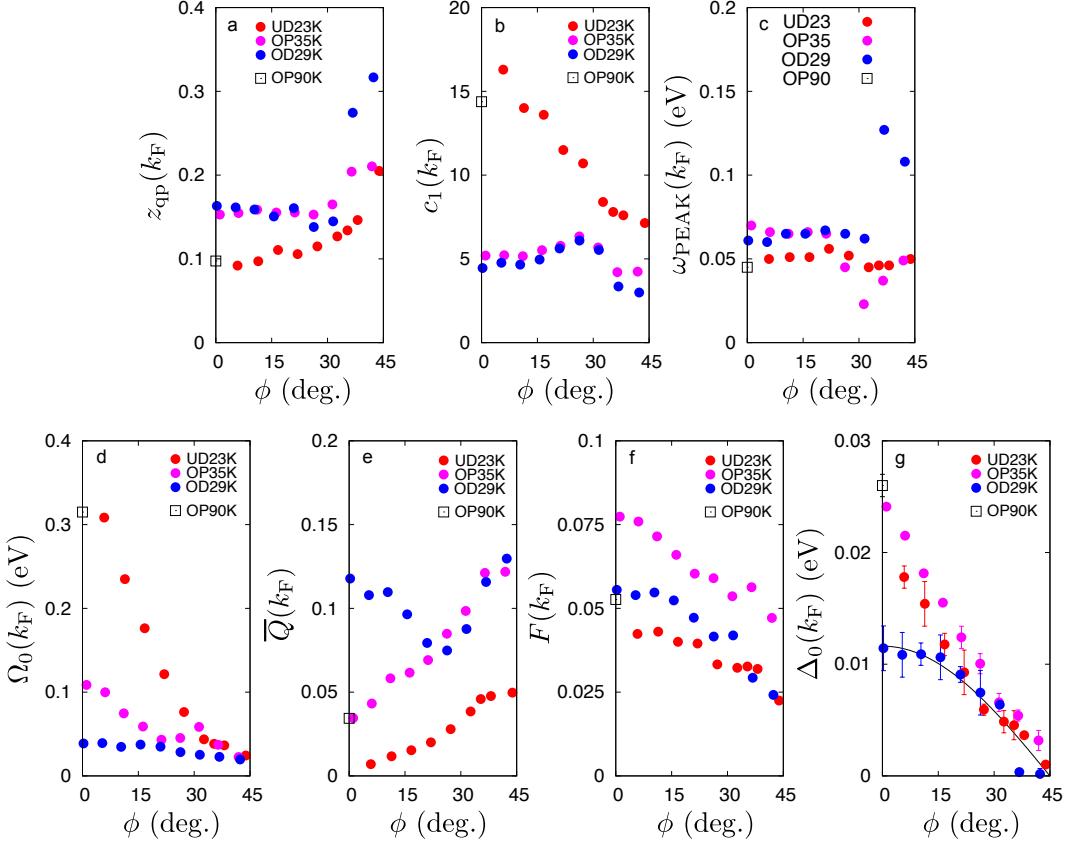


FIG. 17. **Doping and momentum dependences of $z_{qp}(k)$, $c_1(k)$, $\omega_{\text{PEAK}}(k)$, $\overline{W_{\text{PEAK}}}(k)/\omega_{\text{PEAK}}(k)$, $\bar{Q}(k)$, $F(k)$, and $\Delta_0(k)$ for Bi2201 and Bi2212 at $T = 11\text{K}$.** Angle ϕ (taken at Fermi momentum k_F) dependences of a. $z_{qp}(k)$ (Eq.(A7)), b. $c_1(k)$ (Eq. (L1)), c. ω_{PEAK} , d. $\Omega_0(k) = \overline{W_{\text{PEAK}}}(k)/\omega_{\text{PEAK}}$, $\overline{W_{\text{PEAK}}}(k) \equiv \int d\omega \text{Im}W_{\text{PEAK}}(k, \omega)$, e. $\bar{Q}(k) = \int d\omega \text{Im}W(k, \omega)Q(k, \omega)/\overline{W_{\text{PEAK}}}$, f. $F(k)$ (Eq. (A6)), and g. $\Delta_0(k)$ (Eq. (A8)), for three choices of Bi2201 samples with $T_c \sim 23\text{K}$ (UD), 35K (OP) and 29K(OD) are plotted. Data for Bi2212 with $T_c \sim 90\text{K}$ (OP) at the antinode ($\phi \sim 0$) are also added. The solid curve in the most right panel is a cosine curve fitted to $\Delta_0(k)$ for 29K (OD).

- ¹⁵ T. Maier, D. Poilblanc, D. Scalapino, *Phys. Rev. Lett.* **100**, 237001 (2008).
- ¹⁶ S. Sakai, M. Civelli, M. Imada, *Phys. Rev. Lett.* **116**, 057003 (2016).
- ¹⁷ M. Imada, T. J. Suzuki, *J. Phys. Soc. Jpn.* **88**, 024701 (2019).
- ¹⁸ C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, A. E. Ruckenstein, *Phys. Rev. Lett.* **63**, 1996 (1990).
- ¹⁹ S. Martin, A. T. Fiory, R. M. Fleming, S. L. F., J. V. Waszczak, *Phys. Rev. B* **41**, 846(R) (1990).
- ²⁰ H. Takagi, *et al.*, *Phys. Rev. Lett.* **69**, 2975 (1992).
- ²¹ J. Zaanen, Y.-W. Sun, Y. Liu, S. K., *Holographic duality in condensed matter physics* (Cambridge Univ. Press, 2015).
- ²² J. Zaanen, arXiv:1807.10951.
- ²³ S. A. Hartnoll, A. Lucas, S. Sachdev, *Holographic quantum matter* (2018).
- ²⁴ P. L. Russo, *et al.*, *Phys. Rev. B* **75**, 054511 (2004).
- ²⁵ Y. Ando, S. Komiya, K. Segawa, S. Ono, Y. Kurita, *Phys. Rev. Lett.* **93**, 267001 (2004).
- ²⁶ C. C. Homes, *et al.*, *Nature* **430**, 539 (2004).
- ²⁷ Y. J. Uemura, *et al.*, *Phys. Rev. Lett.* **62**, 2317 (1989).
- ²⁸ D. L. Feng, *et al.*, *Science* **289**, 277 (2000).
- ²⁹ H. Ding, *et al.*, *Phys. Rev. Lett.* **87**, 227001 (2001).
- ³⁰ J. S. Kim, G. N. Tam, G. R. Stewart, *Phys. Rev. B* **92**, 224509 (2015).
- ³¹ S. Theodoridis, *Machine Learning - A Bayesian and Optimization Perspective* (Academic Press, Elsevier, Amsterdam, 2015).
- ³² G. Carleo, M. Troyer, *Science* **355**, 602 (2017).
- ³³ N. Le Roux, Y. Bengio, *Neural Comput.* **20**, 1631 (2008).
- ³⁴ P. Morel, P. W. Anderson, *Phys. Rev.* **125**, 1263 (1962).
- ³⁵ K. Hornik, M. Stinchcombe, H. White, *Neural Networks* **2**, 359 (1989).
- ³⁶ S.-I. Amari, *Neural Comput.* **10**, 251 (1998).
- ³⁷ S. Sorella, *Phys. Rev. Lett.* **80**, 4558 (1998).
- ³⁸ S. Sorella, *Phys. Rev. B* **64**, 024512 (2001).
- ³⁹ C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
- ⁴⁰ H. Akaike, *Selected Papers of Hirotugu Akaike* (Springer, 1998), pp. 199–213.
- ⁴¹ J. Bardeen, L. N. Cooper, J. R. Schrieffer, *Physical Review* **108**, 1175 (1957).
- ⁴² H. Li, *et al.*, *Nat. Commun.* **9**, 26 (2018).

TABLE I. Doping dependence of superconducting order parameters and quasiparticle gap. Physical quantities are calculated by the self-energy inferred from ARPES data of three Bi2201 samples and Bi2212 at the optimum doping¹⁰. The doping dependence of the superconducting order parameter $F(k)$, gap amplitude determined from the peak position in EDC, the superconducting gap amplitude $\Delta_0(k)$ obtained by the present Boltzmann machine learning (BML), the quasiparticle renormalization factor z_{qp} , $c_1(k)$, \bar{Q} , $\overline{W}_{\text{PEAK}}$, ω_{PEAK} and Ω_0 are shown mainly at $k = k_{\text{AN}}$, where the doping p is estimated by doping dependence of T_c in⁴⁶. The average over the Fermi surface of F , $\overline{F}(k_F)$ is also estimated.

Sample	Doping		Order Parameter	Gap from EDC		Gap from BML
	T_c (K)	p	$2F(k_{\text{AN}}) \left(\overline{2F}(k_F) \right)$	(meV)	$\Delta_0(k_{\text{AN}})$ (meV)	
Bi2201 UD	23	0.12	0.085 (0.075)	60	18 ± 1	
Bi2201 OP	35	0.15	0.152 (0.126)	30	24.1 ± 0.2	
Bi2201 OD	29	0.18	0.108 (0.089)	12	11 ± 2	
Bi2212 OP	90	-	0.105 (-)	33	26 ± 1	

Sample	Z-factor		Coefficient	Z-factor	Coefficient
	T_c (K)	$z_{\text{qp}}(k_N)$	$c_1(k_N)$	$z_{\text{qp}}(k_{\text{AN}})$	$c_1(k_{\text{AN}})$
Bi2201 UD	23	0.205	7.14	0.092	16.3
Bi2201 OP	35	0.211	4.24	0.155	5.20
Bi2201 OD	29	0.316	3.00	0.16	4.45
Bi2212 OP	90	-	-	0.097	14.4

Sample	Renormalization		Weight	Peak Energy	Energy Scale
	T_c (K)	$\bar{Q}(k_{\text{AN}})$	$\overline{W}_{\text{PEAK}}(k_{\text{AN}})$ (eV ²)	$\omega_{\text{PEAK}}(k_{\text{AN}})$ (eV)	$\Omega_0(k_{\text{AN}})$ (eV)
Bi2201 UD	23	0.007	0.015	0.05	0.31
Bi2201 OP	35	0.035	0.0076	0.07	0.11
Bi2201 OD	29	0.118	0.0024	0.061	0.039
Bi2212 OP	90	0.034	0.014	0.045	0.31

⁴³ D. J. Scalapino, *Superconductivity* (Marcel Dekker Inc., New York, 1969).

⁴⁴ J. R. Schrieffer, *Theory of superconductivity* (CRC Press, 2018).

⁴⁵ M. Le Tacon, *et al.*, *Nat. Phys.* **2**, 537 (2006).

⁴⁶ M. Le Tacon, *et al.*, *Phys. Rev. B* **61**, R14956 (2000).