Quasiparticle interference from magnetic impurities

Philip G. Derry,[1] Andrew K. Mitchell,[1,2] and David E. Logan[1]

[1]*Department of Chemistry, Physical & Theoretical Chemistry, Oxford University, South Parks Road, Oxford OX1 3QZ, United Kingdom*
[2]*Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands*

Fourier transform scanning tunneling spectroscopy (FT-STS) measures the scattering of conduction electrons from impurities and defects, giving information about the electronic structure of both the host material and adsorbed impurities. We interpret such FT-STS measurements in terms of the quasiparticle interference (QPI), here investigating in detail the QPI due to single magnetic impurities adsorbed on a range of representative nonmagnetic host surfaces, and contrasting with the case of a simple scalar impurity or point defect. We demonstrate how the electronic correlations present for magnetic impurities markedly affect the QPI, showing, e.g., a large intensity enhancement due to the Kondo effect, and universality at low temperatures/scanning energies. The commonly used joint density of states interpretation of FT-STS measurements is also considered, and shown to be insufficient in many cases, including that of magnetic impurities.

I. INTRODUCTION

Magnetic atoms embedded on nonmagnetic surfaces provide realizations of quantum impurity models [1] which are amenable to detailed experimental study and manipulation with a scanning tunneling microscope (STM) [2–4]. Such systems are promising candidates as a basis for nanoscale computational, memory storage, and spintronic devices [5,6]. They are also of fundamental interest in their own right, due to the subtle interplay of strongly correlated local spin and orbital degrees of freedom coupled to a conduction electron bath.

In scanning tunneling spectroscopy (STS) experiments, the differential conductance between tip and surface is related to the local density of electronic states (LDOS) at a particular scanning energy and temperature [7,8]; by rastering the STM tip across the surface (Fig. 1) a spatial map of the LDOS in the vicinity of features such as adsorbed impurities may be generated [3]. Such impurities break translational symmetry at the surface, causing scattering of conduction electrons and modulations in the LDOS that depend strongly on both the electronic structure of the underlying sample ("host") and the properties and distribution of impurities. Fourier transform STS (FT-STS), in which such modulations are analyzed in reciprocal space and interpreted in terms of the quasiparticle interference (QPI) between the diagonal states of the clean host [9–11], thus provides a wealth of information on the nature of the impurity-host and (host-mediated) interimpurity correlations.

The aim of this work is to calculate the QPI due to single magnetic adatoms on metallic surfaces, drawing comparison with the case of nonmagnetic, "scalar" impurities (*s*-wave potential scatterers in the weak-scattering Born limit [10]). We also link quantitatively FT-STS measurements to the calculated QPI for generic systems. By way of contrast, we then examine the widely used phenomenological joint density of states (JDOS) interpretation for FT-STS measurements [12,13].

In addition to FT-STS, magnetic impurities can themselves be spectroscopically probed when the STM tip is positioned directly over an impurity. A narrow Kondo resonance is commonly observed around the Fermi energy in these local STS measurements [2,3], arising from the formation of a
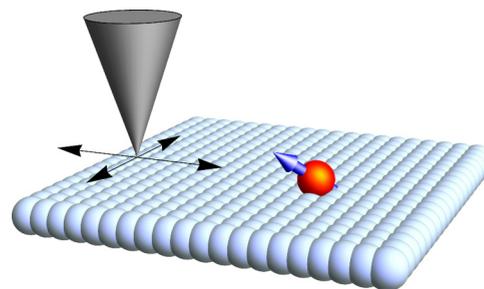
FIG. 1. (Color online) Schematic FT-STS setup: a spatial map of surface LDOS modulations due to electron scattering from a magnetic adatom impurity is extracted from differential conductance measurements as an STM tip is rastered over the surface.

many-body Kondo singlet state where the impurity local moment is screened dynamically by surrounding host conduction electrons. For STS studies of Co adatoms on noble-metal surfaces, typical Kondo temperatures of $T_K \sim 50$–100 K are extracted from the half-width at half-maximum of the spectral Kondo resonance [14]. The single-orbital Anderson impurity model [15] has been successfully used to rationalize such local STS measurements [16–18]. It is adopted here as a qualitatively accurate description of a generic, low-spin magnetic adatom, and is treated using the numerical renormalization group (NRG) [19]. In other cases, e.g., high-spin Mn or Fe adatoms, generalized impurity models must be employed to capture the full orbital structure of the adatom and the host material [20,21] (such calculations are quite specific to the particular system under consideration [22] and are not explicitly considered).

We begin by considering a general formulation (Secs. II and III) for any number or type of impurities embedded on the surface of tight-binding hosts. The single magnetic impurity case, which is the main focus of this paper, is simply a particular example (we subsequently apply the formalism to multi-impurity systems in Ref. [23]). We highlight in particular (Sec. III) the differences between the three-dimensional (3D) cubic lattice with a (100) surface, the two-dimensional (2D) square lattice, and the honeycomb lattice. These simple but

representative hosts reproduce a range of possible material realizations: the LDOS for each features distinct behavior close to the Fermi level, with flat, divergent, and vanishing LDOS for the three lattices, respectively. Each host lattice gives rise to qualitatively different impurity physics with, e.g., the single-impurity Kondo temperature significantly enhanced (suppressed) by increased (depleted) density of states around the Fermi level [24–27].

Single-impurity systems are considered explicitly in Sec. IV and following. The QPI due to single magnetic and scalar impurities for each host is considered in Sec. IV, drawing attention to the qualitative differences in QPI arising from the different types of scattering center and host. The rich *dynamical* properties of the QPI are studied in detail for systems containing a magnetic impurity in Sec. V; we emphasize that this strong energy dependence cannot be reproduced in systems containing only scalar impurities or structural defects. Indeed, strong electron correlations in systems with magnetic impurities are shown to produce unique QPI signatures: in metallic systems, the QPI exhibits universality in terms of rescaled scanning energy and temperature due to the Kondo effect, while nontrivial local-moment physics is observed in the 2D honeycomb case [24,25,28–30].

Section VI examines the interpretation of experimentally measurable FT-STS. By simulating the experimental protocol, we investigate the possible deviation of FT-STS measurements from calculated QPI due to the finite size of the LDOS plaquette measurements in real space. We conclude by critically examining the relationship between the JDOS and the QPI, showing that the two quantities are not directly related in any but the simplest case (a single scalar impurity embedded on a centrosymmetric host), where the JDOS and QPI are Hilbert conjugates.

II. HOST SYSTEMS AND IMPURITY PROBLEM

A. Model

We consider a host material with impurities deposited on the surface, which scatter the quasiparticles of the clean host. Initially, we consider the general problem of N magnetic surface adatoms, formulating the QPI generally for any type, number, and distribution of adatoms.

The clean host is taken to be noninteracting and of tight-binding form, given in its real-space basis by

$$\mathcal{H}_{\text{host}} = \epsilon_0 \sum_{i,\sigma} c_{\mathbf{r}_i\sigma}^\dagger c_{\mathbf{r}_i\sigma} - t \sum_{\langle ij \rangle,\sigma} (c_{\mathbf{r}_i\sigma}^\dagger c_{\mathbf{r}_j\sigma} + \text{H.c.}), \quad (1)$$

where $c_{\mathbf{r}_i\sigma}^\dagger$ creates an electron of spin $\sigma = \uparrow/\downarrow$ in the Wannier orbital localized at site $\mathbf{r}_i$; and $\langle ij \rangle$ denotes a sum over nearest-neighbor sites, coupled by a tunnel matrix element t. Here, we consider the half-filled host $\epsilon_0 = 0$.

Specifically, we focus on the simple 2D square lattice, the 3D cubic lattice with a (100) surface, and the 2D honeycomb lattice. The 3D cubic lattice in particular is representative of a wide class of regular metallic systems, with a constant (finite) electronic density of states at low energies. By contrast, the 2D square lattice, relevant to certain layered materials, features a Van Hove singularity with a logarithmically diverging density of states around the Fermi level [26,31]. The honeycomb

lattice, describing graphene within the simplest noninteracting tight-binding approximation, is notable because it is a bipartite lattice with a vanishing (pseudogapped) density of states at low energies [28,32]. Although simplified, these host systems exemplify a number of distinctive features relevant to real materials, and each induces qualitatively different impurity physics (Sec. IV).

The full model, including N impurities, is given by

$$\mathcal{H} = \mathcal{H}_{\text{host}} + \sum_{\alpha=1}^{N} \mathcal{H}_{\text{imp},\alpha}. \quad (2)$$

In the simplest case, the local potential at site $\mathbf{r}_\alpha$ is modified by impurity α, breaking translational invariance and causing additional electronic scattering. These static defects, referred to as "potential scattering" (or "scalar") impurities, are described by

$$\mathcal{H}_{\text{imp},\alpha}^{ps} = v \sum_{\sigma} c_{\mathbf{r}_\alpha\sigma}^\dagger c_{\mathbf{r}_\alpha\sigma}. \quad (3)$$

This simple model, while often appropriate to describe point defects in materials, does not faithfully capture the physics of many adsorbed impurity adatoms, in particular magnetic impurities, which are *dynamic* objects with internal degrees of freedom and strong local Coulomb interactions [1]. These have a significant qualitative effect on the scattering properties, and must therefore be taken into account. In this paper, we focus primarily on adsorbed magnetic impurities, described in terms of correlated quantum levels

$$\mathcal{H}_{\text{imp},\alpha}^{AIM} = \sum_{\sigma} \epsilon_d d_{\alpha\sigma}^\dagger d_{\alpha\sigma} + U d_{\alpha\uparrow}^\dagger d_{\alpha\uparrow} d_{\alpha\downarrow}^\dagger d_{\alpha\downarrow}$$
$$+ V \sum_{\sigma} (d_{\alpha\sigma}^\dagger c_{\mathbf{r}_\alpha\sigma} + \text{H.c.}), \quad (4)$$

where $d_{\alpha\sigma}^\dagger$ creates a spin-σ electron on impurity α, which is coupled to host site $\mathbf{r}_\alpha$ by tunneling matrix element V. For simplicity, we consider explicitly the particle-hole symmetric case $\epsilon_d = -U/2$. We emphasize that true magnetic impurities preserve SU(2) spin symmetry and time-reversal symmetry, unlike a static polarized local moment, which simply acts like a local magnetic field.

B. Impurity dynamics

Single-particle dynamics of correlated impurities embedded in a noninteracting host are described generically by the Green function matrix $[\mathbf{G}_d(\omega)]_{\alpha,\beta} \equiv G_d^{\alpha\beta}(\omega) = \langle\langle d_{\alpha\sigma}; d_{\beta\sigma}^\dagger \rangle\rangle_\omega$, with $\langle\langle \hat{A}; \hat{B} \rangle\rangle_\omega$ the Fourier transform of the retarded correlator $-i\theta(t)\langle\{\hat{A}(t), \hat{B}(0)\}\rangle$. The matrix Dyson equation for the impurity Green functions is

$$[\mathbf{G}_d(\omega)]^{-1} = [\mathbf{g}_d(\omega)]^{-1} - \mathbf{\Sigma}(\omega). \quad (5)$$

The noninteracting (but host-coupled) impurity propagators are given by

$$[\mathbf{g}_d(\omega)]^{-1} = (\omega + i0^+ - \epsilon_d)\mathbf{I} - \mathbf{\Gamma}(\omega), \quad (6)$$

in terms of the hybridization matrix $\mathbf{\Gamma}(\omega)$ with elements $[\mathbf{\Gamma}(\omega)]_{\alpha,\beta} = V^2 G^0(\mathbf{r}_\alpha, \mathbf{r}_\beta, \omega)$, where $G^0(\mathbf{r}_\alpha, \mathbf{r}_\beta, \omega) = \langle\langle c_{\mathbf{r}_\alpha\sigma}; c_{\mathbf{r}_\beta\sigma}^\dagger \rangle\rangle_\omega^0$ is the propagator between sites $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$

of the clean host (without impurities), and $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ are the host sites to which impurities α and β are coupled. The self-energy matrix $\mathbf{\Sigma}(\omega)$ contains all the nontrivial information due to electronic interactions, which give rise to the Kondo effect, RKKY interaction, etc.

In this work, we employ NRG to solve the underlying quantum impurity problem [19,33]. In the spirit of Ref. [34], equations of motion can be used to obtain an expression for the self-energy matrix

$$\mathbf{\Sigma}(\omega) = [\mathbf{G}_d(\omega)]^{-1}\mathbf{F}_d(\omega), \qquad (7)$$

where $[\mathbf{F}_d(\omega)]_{\alpha,\beta} = U\langle\langle d_{\alpha\sigma}; d^\dagger_{\beta\sigma}d^\dagger_{\beta\bar\sigma}d_{\beta\bar\sigma}\rangle\rangle_\omega$. Both $\mathbf{G}_d(\omega)$ and $\mathbf{F}_d(\omega)$ are calculated directly in NRG using the full density matrix approach [35,36] within the complete Anders-Schiller basis [37].

In the case of a *single* impurity, the Dyson equation [Eq. (5)] reduces to $G^{11}_d(\omega) = [\omega - i0^+ - \epsilon_d - \Gamma(\omega) - \Sigma_{11}]^{-1}$, with the hybridization $\Gamma(\omega) = V^2 G^0(\mathbf{r}_1,\mathbf{r}_1,\omega)$ related simply to the clean host LDOS (itself independent of position due to translational invariance). In NRG [19], a discretized version of the conduction electron Hamiltonian is formulated, and mapped onto a semi-infinite 1D chain with the impurity located at one end. Discretizing on a logarithmic energy scale leads to the "Wilson chain" representation in which hopping matrix elements decrease exponentially down the chain. The RG scheme involves iterative diagonalization, starting at the impurity and working down the Wilson chain, discarding high-energy states at each step [19].

C. Host dynamics

The full dynamics of the host, in the presence of impurities, is embodied in the lattice Green functions $G(\mathbf{r}_i,\mathbf{r}_j,\omega)$ connecting arbitrary host sites $\mathbf{r}_i$ and $\mathbf{r}_j$. They can be related exactly to the above impurity Green functions by equations of motion [38]

$$G(\mathbf{r}_i,\mathbf{r}_j,\omega) - G^0(\mathbf{r}_i,\mathbf{r}_j,\omega)$$
$$= \sum_{\alpha,\beta} G^0(\mathbf{r}_i,\mathbf{r}_\alpha,\omega)T_{\alpha\beta}(\omega)G^0(\mathbf{r}_\beta,\mathbf{r}_j,\omega), \qquad (8)$$

where the sum runs over impurities α and β, and $T_{\alpha\beta}(\omega)$ is the real-space t matrix. For magnetic impurities described by $\mathcal{H}^{AIM}_{\text{imp}}$ in Eq. (4), the t matrix takes the form

$$T^{\text{mag}}_{\alpha\beta}(\omega) = V^2 G^{\alpha\beta}_d(\omega), \qquad (9)$$

requiring as such a knowledge of the full impurity Green functions from Eq. (5). By contrast, the t matrix for potential scattering impurities, described by $\mathcal{H}^{ps}_{\text{imp}}$ in Eq. (3), can be obtained simply in closed form

$$T^{ps}_{\alpha\beta}(\omega) = v[\mathbf{I} - v\mathbf{G}^0(\omega)]^{-1}_{\alpha\beta}, \qquad (10)$$

where the elements $[\mathbf{G}^0(\omega)]_{\alpha'\beta'} \equiv G^0(\mathbf{r}_{\alpha'},\mathbf{r}_{\beta'},\omega)$ are free host Green functions.

In the diagonal quasiparticle basis of the clean host,

$$\mathcal{H}_{\text{host}} = \sum_{\mathbf{k},\sigma} \epsilon_\mathbf{k} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \qquad (11)$$

with $\epsilon_\mathbf{k}$ the dispersion ($\mathbf{k}$ labels the Bloch state momentum). The t-matrix equation [Eq. (8)] can also be transformed into

the momentum-space basis

$$\Delta G(\mathbf{k},\mathbf{k}',\omega) = G^0(\mathbf{k},\omega)T(\mathbf{k},\mathbf{k}',\omega)G^0(\mathbf{k}',\omega), \qquad (12)$$

where $\Delta G(\mathbf{k},\mathbf{k}',\omega) = G(\mathbf{k},\mathbf{k}',\omega) - G^0(\mathbf{k},\omega)\delta_{\mathbf{k},\mathbf{k}'}$ and $G^0(\mathbf{k},\omega) = [\omega + i0^+ - \epsilon_\mathbf{k}]^{-1}$. All quasiparticle scattering induced by the impurities is now contained in

$$T(\mathbf{k},\mathbf{k}',\omega) = \frac{1}{\Omega_{\text{BZ}}} \sum_{\alpha,\beta} e^{i(\mathbf{k}'\cdot\mathbf{r}_\beta - \mathbf{k}\cdot\mathbf{r}_\alpha)} T_{\alpha\beta}(\omega), \qquad (13)$$

where Ω_{BZ} is the volume of the first Brillouin zone.

III. QPI

At sufficiently low temperature and bias, the differential conductance $dI(\mathbf{r},V)/dV$ between STM tip and surface for a particular bias V is proportional to the LDOS, $\rho(\mathbf{r},\omega = eV)$, of the sample at position $\mathbf{r}$. STS thus measures the energy-resolved electronic structure of the sample at a particular point in real space [7,8].

The experimental quantity of interest is the QPI

$$\rho(\mathbf{q},\omega) = \sum_{i\in(L\times L)} e^{-i\mathbf{q}\cdot\mathbf{r}_i} \rho(\mathbf{r}_i,\omega), \qquad (14)$$

obtained at given scanning energy ω. It is therefore the Fourier transform of a real-space LDOS map $\rho(\mathbf{r}_i,\omega)$, itself measured by STS over an $L\times L$ sample region. Since the STS experiment probes the *surface*, the LDOS map is two dimensional. In 2D systems such as graphene, or effective 2D layered systems such as the cuprates, STS thus probes directly the underlying lattice. In 3D systems, by contrast, the lattice probed by STS corresponds to the crystallographic surface lattice.

The FT-STS technique has been employed to investigate the electronic structure of materials and to map their Fermi surface contours [13,39], providing complementary information to techniques such as ARPES [11,12]. FT-STS has proven of particular value in the study of layered materials such as cuprates and pnictides [40–42], as well as topological insulators [43], graphene [44], and heavy-fermion materials [45]. In many cases (e.g., for weak, disorder-induced scattering in cuprates [40]), the local defects giving rise to the QPI may be approximated as scalar impurities in the Born limit [10], although even for scalar impurities this simplification is known to be insufficient in some cases [45]. For, e.g., transition-metal adatoms, however, the full interacting impurity model must be considered.

The QPI at $\mathbf{q} = \mathbf{0}$ is often omitted in experimental results because it corresponds to the *total* density of states sampled, and is extensive in L^2. The desired impurity contribution to the QPI is then obtained by subtracting the result for the clean system without impurities

$$\Delta\rho(\mathbf{q},\omega) = \sum_{i\in(L\times L)} e^{-i\mathbf{q}\cdot\mathbf{r}_i} \Delta\rho(\mathbf{r}_i,\omega), \qquad (15)$$

where $\Delta\rho = \rho - \rho^0$ (with ρ^0 for the clean host). Since $\rho^0(\mathbf{q},\omega) \propto \delta_{\mathbf{q},\mathbf{0}}$, $\Delta\rho(\mathbf{q},\omega)$ scales with the number of impurities. The normalized QPI power spectrum $|\Delta\rho(\mathbf{q},\omega)/N|^2 \propto |\Delta\rho(\mathbf{q},\omega)/L^2|^2$ is *intensive*, independent of the number of impurities or sample region size.

We now give a general formulation for calculating the QPI due to scattering from single or multiple impurities, which can either be simple static potential scattering defects or magnetic (dynamic, interacting) impurities.

A. Real-space formulation

Following the experimental protocol, the QPI can be calculated by discrete Fourier transform of the LDOS within an $L \times L$ region of the host surface, using Eq. (15). The LDOS at site $\mathbf{r}_i$ in the presence of impurities is related to the local host Green function $\rho(\mathbf{r}_i, \omega) = -\frac{1}{\pi} \mathrm{Im} G(\mathbf{r}_i, \mathbf{r}_i, \omega)$, such that from Eq. (8),

$$\Delta \rho(\mathbf{r}_i, \omega) = -\frac{1}{\pi} \mathrm{Im} \sum_{\alpha, \beta} G^0(\mathbf{r}_i, \mathbf{r}_\alpha, \omega) T_{\alpha\beta}(\omega) G^0(\mathbf{r}_\beta, \mathbf{r}_i, \omega) \quad (16)$$

in terms of the full scattering t matrix and free nonlocal host Green functions. The virtue of Eq. (16) is that it is entirely general, and can in principle be used for any lattice with any number or type of impurities. Although the LDOS is *sampled* over a finite region, we stress that the host system is in the thermodynamic limit.

The accurate calculation of lattice Green functions $G^0(\mathbf{r}_i, \mathbf{r}_j, \omega)$ is itself a subtle and well-studied problem [46–49]. Exact diagonalization of finite-sized lattices or discrete Fourier transforms yield poor approximations to Green functions of the desired (semi-)infinite systems, especially at low scanning energies or near Van Hove singularities. Recursion methods [47,50] can be used for periodic systems where the exact dispersion $\epsilon_{\mathbf{k}}$ is known, but such techniques are numerically unstable for large site separations $|\mathbf{r}_i - \mathbf{r}_j|$, and converged solutions can be computationally demanding [51]. Although "bond cutting" [52] and "continued fraction" [53] variants have been developed in special cases, recursion methods typically cannot be used for systems with a surface that breaks translational invariance, limiting applicability within the QPI context. In consequence, calculation of all $N \times L^2$ nonlocal Green functions required for a system of N impurities in an $L \times L$ region is often a major challenge.

In this paper, we have developed instead a technique for fast and accurate numerical calculation of free Green functions on hypercubic-type lattices. The method is detailed in the Appendix, and involves successive convolutions of simpler 1D Green functions which are known exactly in closed form. The convolution itself can moreover be performed efficiently using fast Fourier transform.

B. Scattering state formulation

The "true" QPI containing all scattering information is obtained by taking the thermodynamic limit of plaquette size $L \to \infty$. Translational symmetry implies a basis of states with well-defined momentum *parallel to the surface*, i.e., over the first *surface* Brillouin zone (1SBZ). The local Green function for a surface lattice site is expressed in terms of this basis by 2D Fourier transformation

$$G(\mathbf{r}_i, \mathbf{r}_i, \omega) = \iint_{1SBZ} \frac{d^2\mathbf{k}_\parallel d^2\mathbf{k}'_\parallel}{\Omega_{BZ}} e^{-i\mathbf{r}_i \cdot (\mathbf{k}'_\parallel - \mathbf{k}_\parallel)} G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega), \quad (17)$$

with Ω_{BZ} the volume of the 1SBZ. Writing $\Delta\rho(\mathbf{r}_i, \omega) = -\frac{1}{\pi} \mathrm{Im} \Delta G(\mathbf{r}_i, \mathbf{r}_i, \omega)$, Eq. (15) takes the form

$$\Delta\rho(\mathbf{q}, \omega) = -\frac{1}{\pi} \sum_i e^{-i\mathbf{q}\cdot\mathbf{r}_i} \mathrm{Im} \iint_{1SBZ} \frac{d^2\mathbf{k}_\parallel d^2\mathbf{k}'_\parallel}{\Omega_{BZ}}$$
$$\times e^{-i\mathbf{r}_i \cdot (\mathbf{k}'_\parallel - \mathbf{k}_\parallel)} \Delta G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega), \quad (18)$$

where $\Delta G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega) = G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega) - G^0(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega)$. Equation (18) can itself be recast as

$$\Delta\rho(\mathbf{q}, \omega) = -\frac{1}{2\pi i}[Q(\mathbf{q}, \omega) - Q(-\mathbf{q}, \omega)^*], \quad (19)$$

where

$$Q(\mathbf{q}, \omega) = \int_{1SBZ} d^2\mathbf{k}_\parallel \Delta G(\mathbf{k}_\parallel, \mathbf{k}_\parallel - \mathbf{q}, \omega) \quad (20a)$$

$$\equiv \sum_{\alpha, \beta} T_{\alpha\beta}(\omega) \times \Lambda_{\alpha\beta}(\mathbf{q}, \omega). \quad (20b)$$

As highlighted by Eq. (20b), the QPI factorizes into a momentum-independent scattering amplitude $T_{\alpha\beta}(\omega)$ [Eqs. (9) or (10)], and a host response function $\Lambda_{\alpha\beta}(\mathbf{q}, \omega)$ which depends only on the host lattice and the spatial location of impurities, but not the type of impurity (and thus details of the scattering). The explicit form of this host function must be determined separately for each lattice, as considered in the following.

1. 2D square lattice

Consider first the 2D square lattice, where the QPI calculation is simplest. As the system is itself two dimensional, the surface-momentum basis is simply the diagonal representation $\mathbf{k} = \mathbf{k}_\parallel$. The QPI thus follows from Eqs. (20a), (12), and (13), and is indeed of form Eq. (20b) with

$$\Lambda_{\alpha\beta}(\mathbf{q}, \omega) = \int_{1SBZ} \frac{d^2\mathbf{k}}{\Omega_{BZ}} G^0(\mathbf{k}, \omega) G^0(\mathbf{k} - \mathbf{q}, \omega) e^{i[\mathbf{k}\cdot\mathbf{r}_\alpha - (\mathbf{k} - \mathbf{q})\cdot\mathbf{r}_\beta]}. \quad (21)$$

The free momentum-space Green functions are themselves given by [54] $G^0(\mathbf{k}, \omega) = [\omega + i0^+ - \epsilon_{\mathbf{k}}]^{-1}$, with 2D square lattice dispersion (and lattice constant a_0)

$$\epsilon_{\mathbf{k}} = -2t[\cos(a_0 k_x) + \cos(a_0 k_y)]. \quad (22)$$

The half-bandwidth is then $D = 4t$ in terms of the lattice hopping matrix element t appearing in Eq. (1).

$\Lambda_{\alpha\beta}(\mathbf{q}, \omega)$ can be computed efficiently by using the convolution theorem to do the Brillouin zone integration:

$$\mathcal{F}_{\mathbf{k}}[\Lambda_{\alpha\beta}(\mathbf{q}, \omega)] = \mathcal{F}_{\mathbf{k}}[G^0(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{r}_\alpha}] \times \mathcal{F}_{\mathbf{k}}[G^0(\mathbf{k}, \omega) e^{-i\mathbf{k}\cdot\mathbf{r}_\beta}],$$

where $\mathcal{F}_{\mathbf{k}}$ denotes the 2D fast Fourier transform.

2. 3D cubic lattice with (100) surface

Three-dimensional host lattices are more subtle, due to the surface-sensitive STM measurement. As only the surface LDOS is probed, the QPI amounts to a partial trace over the t-matrix equation (12), in contrast to the full trace for the 2D square lattice [Eq. (21)]. The QPI must thus be evaluated in a basis which preserves the layer index, the

surface momentum basis of Eq. (17) [rather than the diagonal basis of Eq. (12)]. $\Delta G(\mathbf{k}_\parallel, \mathbf{k}_\parallel - \mathbf{q}, \omega)$ in Eq. (20) thus involves propagators between states with surface momentum $\mathbf{k}_\parallel$ and $\mathbf{k}_\parallel - \mathbf{q}$. The 2D transform of $G(\mathbf{r}_i, \mathbf{r}_i, \omega)$ [Eq. (17)] leads to a diagonal representation in each 2D plane in isolation, but surface states labeled by $\mathbf{k}_\parallel$ remain coupled to the bulk (and thus to each other). In general, $\Delta G(\mathbf{k}_\parallel, \mathbf{k}_\parallel - \mathbf{q}, \omega)$ does not therefore take the form of Eq. (12), but rather

$$\Delta G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega) = \iint_{\mathrm{1SBZ}} \frac{d^2\mathbf{k}''_\parallel d^2\mathbf{k}'''_\parallel}{\Omega_{\mathrm{BZ}}} \tilde{G}^0(\mathbf{k}_\parallel, \mathbf{k}''_\parallel, \omega)$$
$$\times\, T(\mathbf{k}''_\parallel, \mathbf{k}'''_\parallel, \omega) \tilde{G}^0(\mathbf{k}'''_\parallel, \mathbf{k}'_\parallel, \omega), \quad (23)$$

where $\tilde{G}^0(\mathbf{k}_\parallel, \mathbf{k}''_\parallel, \omega)$ is a complex through-bulk propagator. Calculation of the QPI thus in general requires the full integrals over intermediate scattering pathways.

In the case of hypercubic-type lattices, however, a significant simplification arises because $\mathbf{k}_\parallel$ is still a good quantum number, so a surface-momentum t-matrix equation with the same structure as Eq. (12) still applies, albeit with modified host surface Green functions. The result for the 3D cubic lattice with a (100) surface is simply

$$\Delta G(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega) = G^0_{\mathrm{surf}}(\mathbf{k}_\parallel, \omega) T(\mathbf{k}_\parallel, \mathbf{k}'_\parallel, \omega) G^0_{\mathrm{surf}}(\mathbf{k}'_\parallel, \omega), \quad (24)$$

where the t matrix is still given by Eq. (13), but

$$G^0_{\mathrm{surf}}(\mathbf{k}_\parallel, \omega) = f\left(\frac{\omega - \epsilon_{\mathbf{k}_\parallel}}{2t}\right), \quad \text{where}$$
$$tf(\tilde{\omega}) = \tilde{\omega} - \begin{cases} \mathrm{sgn}(\tilde{\omega})\sqrt{\tilde{\omega}^2 - 1}, & |\tilde{\omega}| > 1 \\ i\sqrt{1 - \tilde{\omega}^2}, & |\tilde{\omega}| \leqslant 1 \end{cases} \quad (25)$$

with $\epsilon_{\mathbf{k}_\parallel}$ the *2D square lattice* dispersion [Eq. (22)]. This broadens the pole in $G^0(\mathbf{k}, \omega)$ arising for the pure 2D system, to an ellipse of width $2t$ centered on $\omega = \epsilon_{\mathbf{k}_\parallel}$ in $G^0_{\mathrm{surf}}(\mathbf{k}_\parallel, \omega)$ for the 3D system. It follows that the structure of the host function $\Lambda_{\alpha\beta}(\mathbf{q}, \omega)$ is the same as in the 2D square case [Eq. (21)], with $G^0_{\mathrm{surf}}(\mathbf{k}_{||}, \omega)$ in place of $G^0(\mathbf{k}, \omega)$. In this paper, we calculate the true 3D cubic QPI via Eqs. (19), (20), and (21), using the exact expression for the bulk-coupled surface Green functions (25).

For more complex systems where such a prescription is not available, the bulk dephasing of pure 2D surface states could be approximated by using $G^0(\mathbf{k}_\parallel, \omega) = [\omega + i\eta - \epsilon_{\mathbf{k}_\parallel}]^{-1}$, with finite $\eta > 0$. Green function poles are thereby lifetime broadened by Lorentzians of width η.

3. Honeycomb lattice

The honeycomb lattice is complicated by the bipartite nature of the lattice, which is viewed as two interlocking triangular sublattices. We define

$$t(\mathbf{k}) = t e^{-i\mathbf{k}\cdot\boldsymbol{\tau}}[1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2}] \quad (26)$$

in terms of the triangular sublattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and the intersublattice vector $\boldsymbol{\tau} = \mathbf{r}_i^A - \mathbf{r}_i^B$ (where $\mathbf{r}_i^\gamma$ is a site i on the $\gamma = A/B$ sublattice). The A/B-sublattice structure gives rise to distinct $+/-$ bands in momentum space [32]. The honeycomb lattice dispersion for these $+/-$ bands, and the

complex phase, follow as

$$\epsilon_{\mathbf{k}}^\pm = \pm|t(\mathbf{k})|, \quad (27a)$$

$$\phi(\mathbf{k}) = \arg[t(\mathbf{k})]. \quad (27b)$$

Real-space operators are expressed in a diagonal basis by

$$c_{\mathbf{r}_i^\gamma} = \frac{1}{\sqrt{2}} \int_{\mathrm{1BZ}} \frac{d^2\mathbf{k}}{\Omega_{\mathrm{BZ}}^{1/2}} e^{i\mathbf{r}_i^\gamma \cdot \mathbf{k}} e^{i s_\gamma \phi(\mathbf{k})/2} [c_{-,\mathbf{k}} + s_\gamma c_{+,\mathbf{k}}], \quad (28)$$

where $s_\gamma = \pm1$ for the A or B sublattice.

Generalizing Eqs. (15) and (16) to take account of this sublattice structure gives

$$\Delta\rho(\mathbf{q}, \omega) = \sum_{\gamma, \gamma_1, \gamma_2} \Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega), \quad (29)$$

where

$$\Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega) = -\frac{1}{\pi} \sum_{i \in \gamma} e^{-i\mathbf{q}\cdot\mathbf{r}_i^\gamma} \mathrm{Im} \sum_{\alpha \in \gamma_1} \sum_{\beta \in \gamma_2} G^0(\mathbf{r}_i^\gamma, \mathbf{r}_\alpha^{\gamma_1}, \omega)$$
$$\times\, T_{\alpha\beta}(\omega) G^0(\mathbf{r}_\beta^{\gamma_2}, \mathbf{r}_i^\gamma, \omega) \quad (30)$$

with the real-space sum over i spanning sites $\mathbf{r}_i^\gamma$ on sublattice γ. Impurity α (β) is taken to be on sublattice γ_1 (γ_2). Thus, $\Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega)$ is the contribution to the full QPI from sites on the γ sublattice due to impurity-induced scattering between γ_1 and γ_2 sublattices.

Using Eq. (28) in the definition $G^0(\mathbf{r}_i^\gamma, \mathbf{r}_j^{\gamma'}, \omega) = \langle\langle c_{\mathbf{r}_i^\gamma}; c_{\mathbf{r}_j^{\gamma'}}^\dagger \rangle\rangle_\omega^0$, $\Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega)$ takes the same form as Eq. (18):

$$\Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega) = -\frac{1}{\pi} \sum_i e^{-i\mathbf{q}\cdot\mathbf{r}_i^\gamma} \mathrm{Im} \iint_{\mathrm{1BZ}} \frac{d^2\mathbf{k}\, d^2\mathbf{k}'}{\Omega_{\mathrm{BZ}}}$$
$$\times\, e^{-i\mathbf{r}_i^\gamma \cdot (\mathbf{k}' - \mathbf{k})} \Delta G_{\gamma, \gamma_1, \gamma_2}(\mathbf{k}, \mathbf{k}', \omega), \quad (31)$$

but now with

$$\Delta G_{\gamma, \gamma_1, \gamma_2}(\mathbf{k}, \mathbf{k}', \omega)$$
$$= \sum_{\alpha \in \gamma_1} \sum_{\beta \in \gamma_2} \frac{T_{\alpha\beta}(\omega)}{4\Omega_{\mathrm{BZ}}} e^{-i(\mathbf{r}_\alpha^{\gamma_1}\cdot\mathbf{k} - \mathbf{r}_\beta^{\gamma_2}\cdot\mathbf{k}')} e^{i[(s_\gamma - s_{\gamma_1})\phi(\mathbf{k}) - (s_\gamma - s_{\gamma_2})\phi(\mathbf{k}')]/2}$$
$$\times [s_\gamma G_-^0(\mathbf{k}) + s_{\gamma_1} G_+^0(\mathbf{k})][s_\gamma G_-^0(\mathbf{k}') + s_{\gamma_2} G_+^0(\mathbf{k}')], \quad (32)$$

in terms of the $+/-$ band free Green functions [54] $G_\pm^0(\mathbf{k}) \equiv \langle\langle c_{\pm,\mathbf{k}}; c_{\pm,\mathbf{k}}^\dagger \rangle\rangle_\omega^0 = [\omega + i0^+ - \epsilon_{\mathbf{k}}^\pm]^{-1}$. The QPI contribution then follows as

$$\Delta\rho_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega) = -\frac{1}{2\pi i}\left[Q_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega) - Q_{\gamma, \gamma_1, \gamma_2}(-\mathbf{q}, \omega)^*\right], \quad (33)$$

where

$$Q_{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega) = \sum_{\alpha \in \gamma_1} \sum_{\beta \in \gamma_2} T_{\alpha\beta}(\omega) \times \Lambda_{\alpha\beta}^{\gamma, \gamma_1, \gamma_2}(\mathbf{q}, \omega). \quad (34)$$

Equations (33) and (34) are thus analogs of Eqs. (19) and (20), with

$$\Lambda_{\alpha\beta}^{\gamma,\gamma_1,\gamma_2}(\mathbf{q},\omega) = \int_{1BZ} \frac{d^2\mathbf{k}}{4\Omega_{BZ}} e^{-i[\mathbf{r}_\alpha^{\gamma_1}\cdot\mathbf{k} - \mathbf{r}_\beta^{\gamma_2}\cdot(\mathbf{k}-\mathbf{q})]}$$
$$\times e^{i[(s_\gamma - s_{\gamma_1})\phi(\mathbf{k}) - (s_\gamma - s_{\gamma_2})\phi(\mathbf{k}-\mathbf{q})]/2}$$
$$\times [s_\gamma G_-^0(\mathbf{k}) + s_{\gamma_1} G_+^0(\mathbf{k})]$$
$$\times [s_\gamma G_-^0(\mathbf{k}-\mathbf{q}) + s_{\gamma_2} G_+^0(\mathbf{k}-\mathbf{q})]. \quad (35)$$

We stress that interband scattering and the momentum-dependent phase factors appearing in Eq. (35) are important, and affect the full QPI qualitatively.

IV. SINGLE-IMPURITY QPI: MAGNETIC AND SCALAR IMPURITIES

The generalized problem involving N magnetic impurities, spatially separated and coupled to conduction electrons of the host lattice, is naturally highly rich and complex (we consider aspects of it in subsequent work [23]). From here on in this paper, we focus on a single magnetic impurity, in terms of which experimental QPI patterns are in fact typically interpreted. In this case, the QPI is given by Eq. (20b) (with $\alpha = \beta = 1$, dropped hereafter):

$$Q(\mathbf{q},\omega) = T(\omega)\Lambda(\mathbf{q},\omega), \quad (36)$$

with $T(\omega)$ the single-impurity t matrix and explicit forms for the host functions $\Lambda(\mathbf{q},\omega)$ given in Secs. IV B–IV D.

The impurity itself is often taken to be a static potential defect in the weak-scattering Born limit [10]. The t matrix is then pure real and energy independent,

$$T^{ps}(\omega) \simeq v, \quad (37)$$

with v the potential scattering strength [see Eq. (3)]. Equation (37) is the leading-order approximation to the exact Eq. (10), holding provided $|vG^0(\mathbf{r}_\alpha, \mathbf{r}_\alpha, \omega)| \ll 1$. We add, however, that this approximation is *not* valid in the vicinity of divergences in the host density of states (arising, e.g., at $\omega = 0$ in the 2D square lattice). In the special case of a single impurity on a centrosymmetric surface $Q(\mathbf{q},\omega) = Q(-\mathbf{q},\omega)$, so [Eq. (19)] $\Delta\rho(\mathbf{q},\omega) = -\frac{1}{\pi}\mathrm{Im}Q(\mathbf{q},\omega)$. For a scalar impurity in the Born limit,

$$\Delta\rho(\mathbf{q},\omega) \overset{\text{scalar}}{=} -\frac{v}{\pi}\Lambda''(\mathbf{q},\omega), \quad (38)$$

where $\Lambda(\mathbf{q},\omega) = \Lambda'(\mathbf{q},\omega) + i\Lambda''(\mathbf{q},\omega)$. The QPI scanning-energy dependence is thus due solely to that of $\Lambda''(\mathbf{q},\omega)$.

For magnetic impurities by contrast, electron correlations give rise to nontrivial dynamics. From Eq. (9),

$$T^{\text{mag}}(\omega) = V^2 G_d(\omega) \quad (39)$$

in terms of the impurity Green function $G_d(\omega)$; the QPI follows from Eqs. (19) and (36). For a centrosymmetric surface,

$$\Delta\rho(\mathbf{q},\omega) \overset{\text{mag}}{=} -\frac{V^2}{\pi}[\mathrm{Re}G_d(\omega)\Lambda''(\mathbf{q},\omega) + \mathrm{Im}G_d(\omega)\Lambda'(\mathbf{q},\omega)], \quad (40)$$

with contributions from both real and imaginary parts of $\Lambda(\mathbf{q},\omega)$, and weights that depend on the impurity Green

function at energy ω. As discussed in the following, the Kondo effect produces a scattering enhancement at low temperatures and scanning energies, causing a crossover in the QPI from being dominated by $\Lambda''(\mathbf{q},\omega)$ at high energies (similar to that of a scalar impurity) to being dominated by $\Lambda'(\mathbf{q},\omega)$ at low energies.

A. Effect of host on impurity dynamics

For a magnetic impurity, the QPI depends on both the host function $\Lambda(\mathbf{q},\omega)$ and the impurity Green function $G_d(\omega)$, which itself depends on the host. Specifically, the impurity problem is controlled by the hybridization function $\Gamma(\omega)$, related to the clean host density of states $\rho_0(\omega)$, by $-\mathrm{Im}\Gamma(\omega) = \pi V^2 \rho_0(\omega)$. The Kondo physics is sensitive to the behavior of $\rho_0(\omega)$ near the Fermi level ($\omega = 0$) [24–26,30] and to leading order

$$\rho_0(\omega) \overset{|\omega| \ll t}{\sim} \begin{cases} \frac{\ln(16t/|\omega|)}{2\pi^2 t} & : \text{2D square} \\ \frac{1}{6t} - \frac{\omega^2}{6\sqrt{2}\pi^2 t^3} & : \text{3D cubic, (100) surface} \\ \frac{|\omega|}{\sqrt{3}\pi t^2} & : \text{honeycomb}. \end{cases} \quad (41)$$

These lattices exemplify three paradigms, with densities of states that are diverging, flat, or pseudogapped at low energy. $\rho_0(\omega)$ versus ω/t is shown in the insets to Fig. 2.

The density of states for metallic systems is typically flat at low energies. This gives rise to an exponentially small Kondo scale [1] $T_K/t \sim \exp[-\pi U/8V^2\rho_0(0)]$, and low-energy Fermi liquid behavior [1]

$$\mathrm{Im}\Gamma(\omega) \times \mathrm{Im}G_d(\omega) \overset{|\omega| \ll T_K}{\sim} 1 - \alpha_\omega(\omega/T_K)^2 + \cdots, \quad (42)$$

with Fermi level spectral pinning $-\pi V^2 \rho_0(0)\mathrm{Im}G_d(0) = 1$ [as $\mathrm{Im}\Gamma(0) = -\pi V^2 \rho_0(0)$]. This is shown for a magnetic impurity on the (100) surface of a 3D cubic lattice in Fig. 2 (center panel), where the imaginary and real parts of $tG_d(\omega)$ are plotted versus ω/t. We have chosen representative impurity parameters $U = 1.95t$ and $V = 0.56t$, yielding $T_K \approx 5 \times 10^{-3}t$ (defined here as the half-width at half-maximum of the Kondo resonance [55]). With the host bandwidth $12t = 11$ eV (such that $U = 1.79$ eV and $V = 0.51$ eV) we obtain $T_K \approx 57$ K, consistent with established results [14,17,56] for Co atoms on a Cu surface. As seen from Fig. 2 (center), the Kondo effect results in a large imaginary part $-t\mathrm{Im}G_d(\omega) \sim 6t^2/(\pi V^2) \simeq 6.2$ for low energies $|\omega| \ll T_K$.

For the 2D square lattice, the low-energy divergence in the host density of states results in an enhanced Kondo temperature [26]. In any Kondo phase, the pinning condition from Eq. (42) still holds [30,57], implying that

$$-\mathrm{Im}G_d(\omega) \overset{|\omega| \to 0}{\sim} \frac{2\pi t}{V^2}[\ln(16t/|\omega|)]^{-1}, \quad (43)$$

which decays logarithmically at low energies. As confirmed in the left panel of Fig. 2, the impurity spectrum $-\mathrm{Im}tG_d(\omega)$ therefore shows a maximum at $|\omega| \sim T_K$. With the same parameters as the 3D cubic system, we now obtain a much higher $T_K \approx 584$ K [55].

Finally, in the pseudogapped honeycomb lattice the Kondo effect is suppressed due to the depleted density of states near the Fermi level, and the local moment phase is stable for any U/V^2 at particle-hole symmetry [24,28]. The impurity
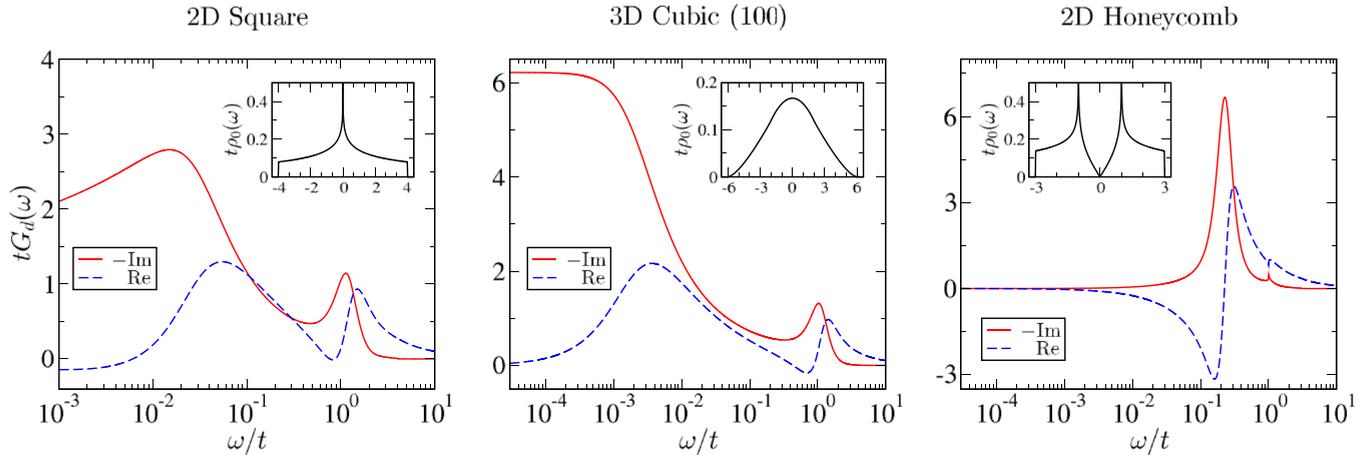
FIG. 2. (Color online) Dynamics of a single magnetic impurity on various lattices at $T = 0$, plotted as $-\mathrm{Im}\, t G_d(\omega)$ (solid lines) and $\mathrm{Re}\, t G_d(\omega)$ (dashed) vs ω/t, calculated via NRG. Insets show corresponding host density of states $t\rho_0(\omega)$. Impurity parameters: $U = 1.95t$ and $V = 0.555t$ for 2D square and 3D cubic lattices; $U = 0.704t$ and $V = 0.493t$ for honeycomb lattice.

spectrum then takes the low-energy form

$$- \mathrm{Im}\, G_d(\omega) \overset{|\omega| \to 0}{\sim} \alpha' |\omega| \qquad (44)$$

(with α' a constant). This decay of $G_d(\omega)$ is demonstrated in the right panel of Fig. 2. With $t = 2.84$ eV, the honeycomb lattice models the π/π^* bands of graphene [58]; we use $U = 0.704t$ and $V = 0.493t$ as realistic impurity parameters obtained from *ab initio* calculations for Co atoms on graphene [59].

We now turn to the QPI for these three lattices.

B. 2D square lattice

The QPI is obtained from Eq. (38) or (40), with $\Lambda(\mathbf{q},\omega)$ for a single impurity given from Eq. (21) (with $\mathbf{r}_\alpha = \mathbf{r}_\beta = \mathbf{0}$):

$$\Lambda(\mathbf{q},\omega) = \int_{1\mathrm{SBZ}} \frac{d^2\mathbf{k}}{\Omega_{\mathrm{BZ}}} G^0(\mathbf{k},\omega) G^0(\mathbf{k}-\mathbf{q},\omega). \qquad (45)$$

Figure 3 shows the absolute value of the QPI $|\Delta\rho(\mathbf{q})|$ as a color map in $\mathbf{q}$ space (upper panels), comparing the scalar impurity (left) with the magnetic impurity (right), at a fixed scanning energy $\omega \simeq T_K$, using the same parameters as Fig. 2. The lower panel shows a cut across the Brillouin zone of $\Delta\rho(\mathbf{q})/\Delta\rho_{\mathrm{tot}}$, where $\Delta\rho_{\mathrm{tot}} \equiv \Delta\rho_{\mathrm{tot}}(\omega) = \int_{1\mathrm{BZ}} d^2q |\Delta\rho(\mathbf{q},\omega)|$ is the total scattering amplitude at energy ω. For a single impurity, the topology in $\mathbf{q}$ space, on which we now focus, is completely determined by the host function $\Lambda(\mathbf{q},\omega)$ [see Eqs. (38) and (40)].

For scalar impurities in the Born limit, only the imaginary part $\Lambda''(\mathbf{q},\omega)$ plays a role [Eq. (38)]. For given scanning energy ω, its structure gives rise to singular lines at $\mathbf{q} = \mathbf{q}^*(\omega)$ associated with the Van Hove point of the 2D square lattice. Around the Γ symmetry point, these lines are rectangular hyperbola

$$(q_x^*)^2 - (q_y^*)^2 \overset{|\omega| \ll t}{\sim} \pm a_\Gamma(\omega)^2 \ : -\pi < q_{x,y}^* \leqslant \pi \qquad (46)$$

with the dispersive properties controlled by $a_\Gamma(\omega)$, while around the M point

$$(q_x^* \pm q_y^*)^2 \overset{|\omega| \ll t}{\sim} [2\pi - a_M(\omega)]^2 \ : -\pi < q_{x,y}^* \leqslant \pi. \qquad (47)$$

In $\mathbf{q}$ space, $\Lambda''(\mathbf{q},\omega)$ is found to diverge as

$$\Lambda''(\mathbf{q},\omega) \overset{\mathbf{q} \to \mathbf{q}^*}{\sim} |\mathbf{q} - \mathbf{q}^*|^{-1/2} \qquad (48)$$

when approaching a point $\mathbf{q}^*$ from the Γ or M points. The region enclosed by these divergences is therefore characterized by high QPI scattering intensity (see Fig. 3 for the scalar impurity). $\Lambda''(\mathbf{q},\omega)$ does not, however, diverge on approaching from X, and remains comparatively small in its vicinity. In fact, $\Lambda''(\mathbf{q},\omega)$ is *odd* in ω due to the exact symmetry $\Lambda(\mathbf{q},\omega) = \Lambda(\mathbf{q}, -\omega)^*$. The scalar impurity QPI $\Delta\rho(\mathbf{q},\omega) \sim \omega$ thus vanishes at low energies away from the lines of divergence.

The situation is rather different for the magnetic impurity because both real and imaginary parts of $\Lambda(\mathbf{q},\omega)$ are important [Eq. (40)]. Because $\Lambda'(\mathbf{q},\omega)$ is even in ω, residual QPI intensity around the X symmetry point persists even at low energies, due to finite $\Lambda'(\mathbf{q},\omega=0) = b'_\mathbf{q}$. $\Lambda'(\mathbf{q},\omega)$ also diverges logarithmically [as Eq. (48)] on approaching the singular lines from X (it does not diverge in the vicinity of Γ or M). As such, the QPI scattering intensity is enhanced around X for magnetic impurities.

Further, as shown in the lower panel of Fig. 3, the *sign* of $\Delta\rho(\mathbf{q},\omega)$ can be reversed by the contribution from the second term in Eq. (40). This is a hallmark of scattering from magnetic impurities, where $\mathrm{Im}\, G_d(\omega) < 0$ can become large due to the Kondo effect (see Fig. 2). This leads to additional structure in the measurable $|\Delta\rho(\mathbf{q})|$, not found in QPI for scalar impurities.

C. 3D cubic lattice with (100) surface

The (100) surface of the 3D cubic lattice is again a square lattice, but "surface" quasiparticles are dephased by coupling to the bulk. This leads to the t matrix (24), and for a single impurity

$$\Lambda(\mathbf{q},\omega) = \int_{1\mathrm{SBZ}} \frac{d^2\mathbf{k}_\parallel}{\Omega_{\mathrm{BZ}}} G^0_{\mathrm{surf}}(\mathbf{k}_\parallel,\omega) G^0_{\mathrm{surf}}(\mathbf{k}_\parallel - \mathbf{q},\omega) \qquad (49)$$

with scattering vectors $\mathbf{q} \equiv \mathbf{q}_\parallel$ confined to the 2D surface [and the $G^0_{\mathrm{surf}}(\mathbf{k}_\parallel,\omega)$ given by Eq. (25)].

The resulting QPI, shown in Fig. 4, does not contain the divergences arising in the 2D square lattice, but remnants of
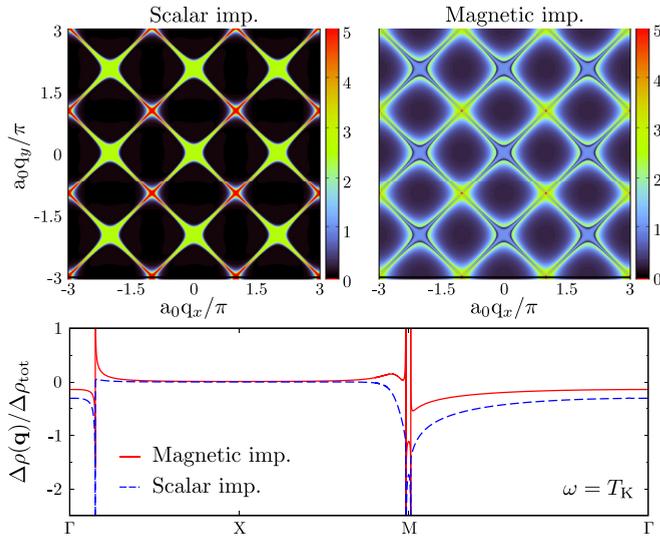
FIG. 3. (Color online) QPI for a single impurity on the 2D square lattice at a scanning energy $\omega = 0.055t \simeq T_K$. Upper panels compare the QPI maps $|\Delta\rho(\mathbf{q})|$ for scalar (left) and magnetic (right) impurities; lower panel shows a Brillouin zone cut along the path $\Gamma \to \mathrm{X} \to \mathrm{M} \to \Gamma$, plotted as $\Delta\rho(\mathbf{q})/\Delta\rho_{\mathrm{tot}}$. Symmetry points defined as $\mathbf{q}_\Gamma = \mathbf{0}$, $\mathbf{q}_{\mathrm{X}} = \mathbf{A}_1$, $\mathbf{q}_{\mathrm{M}} = \mathbf{A}_1 + \mathbf{A}_2$ in terms of reciprocal lattice vectors $\mathbf{A}_1 = 2\pi/a_0(1,0)$, $\mathbf{A}_2 = 2\pi/a_0(0,1)$. Magnetic impurity parameters as in Fig. 2 such that $\omega = 5\times10^{-2}t = T_K$; and $v = 0.5t$ for the scalar impurity.

this singular structure appear in broadened regions of enhanced scattering intensity around the M symmetry point in the cubic lattice. The global fourfold symmetry of the QPI evolves locally into a continuous rotational symmetry around this point, with $\Lambda(\mathbf{q},\omega) \equiv \Lambda(|\mathbf{q} - \mathbf{q}_{\mathrm{M}}|,\omega)$ for $|\omega| \ll t$. Further, the QPI for the scalar impurity [which only depends on $\Lambda''(\mathbf{q},\omega)$] is distinctly conical, with

$$\Lambda''(\mathbf{q},\omega) \stackrel{\mathbf{q}\to\mathbf{q}_{\mathrm{M}}}{\sim} a''_{\mathrm{M}}(\omega) + b''_{\mathrm{M}}(\omega)|\mathbf{q} - \mathbf{q}_{\mathrm{M}}| + \mathcal{O}|\mathbf{q} - \mathbf{q}_{\mathrm{M}}|^2, \quad (50)$$

whereas the QPI for a magnetic impurity can become dominated by the quadratic term (see Fig. 4) since

$$\Lambda'(\mathbf{q},\omega) \stackrel{\mathbf{q}\to\mathbf{q}_{\mathrm{M}}}{\sim} a'_{\mathrm{M}}(\omega) + c'_{\mathrm{M}}(\omega)|\mathbf{q} - \mathbf{q}_{\mathrm{M}}|^2 + \mathcal{O}|\mathbf{q} - \mathbf{q}_{\mathrm{M}}|^4. \quad (51)$$

A striking feature of the QPI for the 3D cubic lattices is the difference in *intensity* between scalar and magnetic impurities (note the rescaled color range in Fig. 4). There are two distinct reasons for this. First, $\Lambda''(\mathbf{q},\omega)$ is odd in ω, whence $\Lambda''(\mathbf{q},\omega) \sim \omega$ at low energies $|\omega| \ll t$. For the scalar impurity, Eq. (38) implies that the QPI, $\Delta\rho(\mathbf{q}) \sim \omega$, is therefore also small. By contrast, the QPI for a magnetic impurity [Eq. (40)] has a contribution from $\Lambda'(\mathbf{q},\omega)$, which remains finite as $\omega \to 0$.

Second, in the case of magnetic impurities, the Kondo effect produces a spectral resonance in $\mathrm{Im}G_d(\omega)$ of width T_K that does not decay at low energies [cf. Eq. (43) and Fig. 2]. In consequence, the QPI is considerably more intense at low energies for magnetic impurities than scalar impurities in standard flat-band metallic systems.
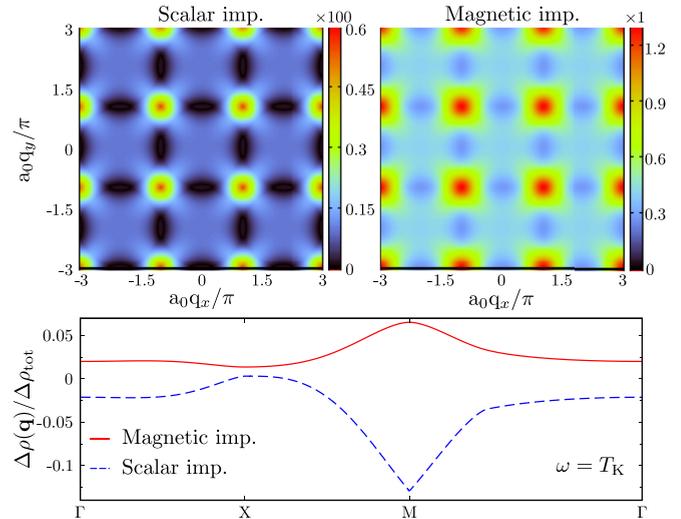
FIG. 4. (Color online) QPI for an impurity on the (100) surface of a 3D cubic lattice. As Fig. 3 but at $\omega = 5\times10^{-3}t = T_K$.

D. 2D honeycomb lattice

The 2D honeycomb lattice, modeling the π and π^* bands in graphene within a nearest-neighbor tight-binding picture [58], generates richer structure in the QPI [13,59,60] due to the bipartite nature of the lattice and the low-energy pseudogapped density of states [Eq. (41)].

A single impurity coupled to a single honeycomb site (on sublattice $\gamma' = A$ or B) lowers the symmetry by breaking the centrosymmetry of the lattice. The single-impurity QPI is therefore obtained from Eqs. (19) and (36), with $\Lambda(\mathbf{q},\omega)$ comprising contributions from both sublattices

$$\Lambda(\mathbf{q},\omega) = \sum_\gamma \Lambda_{\alpha\alpha}^{\gamma,\gamma',\gamma'}(\mathbf{q},\omega). \quad (52)$$

$\Lambda_{\alpha\alpha}^{\gamma,\gamma',\gamma'}(\mathbf{q},\omega)$ itself is given by Eq. (35), and depends on the phase ϕ defined in Eq. (27b). This phase has a marked qualitative effect on the resulting QPI, and cannot be neglected. For centrosymmetric lattices, the QPI is periodic across the first Brillouin zone because $\Lambda(\mathbf{q} + n\mathbf{A}_i,\omega) = \Lambda(\mathbf{q},\omega)$ with integer n for any reciprocal lattice vector $\mathbf{A}_i$. But, in the honeycomb lattice

$$\Lambda(\mathbf{q} + 3n\mathbf{A}_i,\omega) = \Lambda(\mathbf{q},\omega), \quad (53)$$

arising because $\phi(\mathbf{k} + n\mathbf{A}_i) = \phi(\mathbf{k}) + \exp[2n\pi i/3]$ [61]. As such, the period of the QPI is enlarged to include the third Brillouin zone.

One consequence of this is the inequivalence of intravalley scattering at the Γ and Γ_2 points (located at $\mathbf{q}_\Gamma = \mathbf{0}$ and $\mathbf{q}_{\Gamma_2} = \mathbf{A}_1$, respectively), where $\Delta\rho(\mathbf{q}_{\Gamma_2},\omega) = \Delta\rho(\mathbf{q}_\Gamma,\omega)\times \exp[-\pi i/3]$. At low energies $|\omega| \ll t$, these points are surrounded by singular lines in $\Lambda(\mathbf{q},\omega)$ at $\mathbf{q} = \mathbf{q}^*$, where

$$|\mathbf{q} - \mathbf{q}^*| = d_\Gamma(\omega), \quad (54)$$

giving rise to a circular derivative discontinuity in the QPI around $\mathbf{q}_\Gamma$ and $\mathbf{q}_{\Gamma_2}$. The dispersive properties are controlled by $d_\Gamma(\omega) = d_{\Gamma_2}(\omega)$, discussed further in the following subsection. For the scalar impurity at low energies $|\omega| \ll t$, both points

are surrounded by flat regions of high scattering intensity

$$|\Delta\rho(\mathbf{q},\omega)| \overset{|\mathbf{q}|<|\mathbf{q}^*|}{\sim} b_\Gamma \qquad (55)$$

with $b_\Gamma = b_{\Gamma_2}$ independent of scanning energy ω. However, the local environment of the Γ and Γ_2 points is different. The immediate vicinity of the Γ_2 point possesses a continuous rotational symmetry, with divergences in the QPI along the entire singular line $\mathbf{q} = \mathbf{q}^*$:

$$\Lambda(\mathbf{q},\omega) \overset{\mathbf{q}\to\mathbf{q}^*}{\sim} \ln^2 |a_0\mathbf{q} - a_0\mathbf{q}^*|. \qquad (56)$$

By contrast, a lower sixfold symmetry is found around the Γ point as $\omega \to 0$ due to divergent *points* arising only when $\mathbf{q}^* \times \boldsymbol{\delta} = \mathbf{0}$, with $\boldsymbol{\delta} = \mathbf{A}_1, \mathbf{A}_2,$ and $(\mathbf{A}_1 + \mathbf{A}_2)$.

Complex features in the QPI also appear in the vicinity of the K symmetry points due to intervalley scattering, and are again enclosed by singular lines, denoted $\mathbf{q}^*$. These features possess only reflection symmetry about the line $(\mathbf{q}_\Gamma - \mathbf{q}_K)$, the continuous rotational symmetry being lifted by the underlying phase texture (itself arising because the impurity couples to a single sublattice). The line of divergence along $\mathbf{q}^*$ is intersected by a perpendicular nodal line at pinch points where $(\mathbf{q}^* - \mathbf{q}_K) \cdot (\mathbf{q}_\Gamma - \mathbf{q}_K) = 0$. For the scalar impurity, scattering is forbidden within the region around the K point enclosed by the singular lines. These features are seen clearly in the QPI map and cuts for the scalar impurity presented in Fig. 5. At higher scanning energies, trigonal warping sets in, giving rise to a local threefold point symmetry around $\mathbf{q}_K$.

For the magnetic impurity, the relative weight of $\Lambda'(\mathbf{q},\omega)$ and $\Lambda''(\mathbf{q},\omega)$ in the QPI depends on the complex t matrix $T(\omega)$, which evolves with scanning energy ω (see Fig. 2). Importantly, this can lead to distinctive features in the measurable QPI, $|\Delta\rho(\mathbf{q},\omega)|$. Accidental cancellation of terms in Eq. (19) can produce "dark spots" of suppressed scattering in
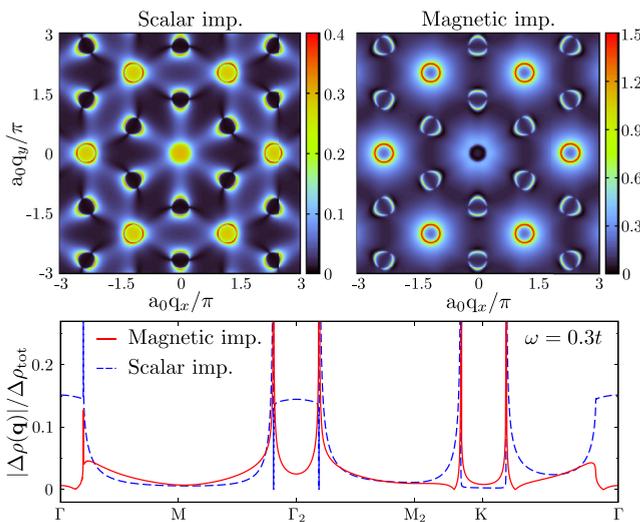
FIG. 5. (Color online) QPI for a single impurity on the honeycomb lattice. As Fig. 3 but at $\omega = 0.3t$ and with impurity parameters from Fig. 2. The Brillouin zone cut takes the path $\Gamma \to M \to \Gamma_2 \to M_2 \to K \to \Gamma$, where $\mathbf{q}_\Gamma = \mathbf{0}$, $\mathbf{q}_M = \frac{1}{2}\mathbf{A}_1$, $\mathbf{q}_{\Gamma_2} = \mathbf{A}_1$, $\mathbf{q}_{M_2} = \mathbf{A}_1 + \frac{1}{2}\mathbf{A}_2$, $\mathbf{q}_K = \frac{2}{3}\mathbf{A}_1 + \frac{1}{3}\mathbf{A}_2$, in terms of reciprocal lattice vectors $\mathbf{A}_{1,2} = 2\pi/a_0(\frac{1}{\sqrt{3}}, \pm 1)$.

the QPI. An example is shown in Fig. 5, where $|\Delta\rho(\mathbf{q},\omega)| \simeq 0$ for $|\mathbf{q}| < |\mathbf{q}^*|$ in the vicinity of the Γ point. In contrast to the scalar impurity case [Eq. (55)], the QPI in general depends on ω and varies with $\mathbf{q}$ in the vicinity of the Γ and Γ_2 points when magnetic impurities are present. Indeed, magnetic impurities also induce scattering near the K point.

V. CHARACTERISTIC KONDO PHYSICS IN THE QPI

A. Scanning-energy dependence

We turn now to dynamical features of the QPI for the three lattices, comparing scalar and magnetic impurities. Numerically exact results which exemplify the key physics are presented in Fig. 6.

For scalar impurities, the scanning-energy dependence of the QPI is due entirely to the ω dependence of $\Lambda(\mathbf{q},\omega)$, which characterizes the clean host lattice. The real part of this function is plotted as a color map in the center column panels of Fig. 6 (the real and imaginary parts are related by Hilbert transformation).

For magnetic impurities, QPI dynamics result from both $\Lambda(\mathbf{q},\omega)$ and the impurity Green function $G_d(\omega)$, whose spectrum is plotted in the left column panels of Fig. 6 [see also Fig. 2 and Eqs. (42)–(44) for the detailed low-energy behavior]. The nontrivial scanning-energy dependence of the QPI reflects the rich structure of the underlying quantum impurity problem.

For the 2D square lattice, divergences in $\Lambda(\mathbf{q},\omega)$, described by Eqs. (46) and (47), give rise to lines of intense scattering in the QPI. The dispersive properties of these features are controlled at low energies $|\omega| \ll t$ by $a_\Gamma(\omega)$ and $a_M(\omega)$, which are related by continuity at the edge of the Brillouin zone through $a_M(\omega) = a_\Gamma(\omega)^2/(2\pi)$. We find linear dispersion of the divergent features around the M symmetry point, implying

$$a_M(\omega) \sim |\omega|; \quad a_\Gamma(\omega) \sim |\omega|^{1/2}, \qquad (57)$$

as confirmed directly in the upper panels of Fig. 6.

As $\omega \to 0$, the divergences are confined to the line $q_x^* = q_y^*$ connecting Γ and M symmetry points, whence

$$\Lambda(q_x = q_y, \omega) \overset{|\omega|\ll t}{\sim} a_\mathbf{q}\delta(\omega) + b_\mathbf{q}|\omega| + ic_\mathbf{q}\frac{1}{\omega} + \cdots. \qquad (58)$$

The QPI itself thus diverges along this line with the universal asymptotic form

$$|\Delta\rho(q_x = q_y, \omega)| \overset{|\omega|\to 0}{\sim} \begin{cases} \left|\frac{1}{\omega}\right| & : \text{scalar}, \\ \frac{1}{|\omega|\ln^2(16t/|\omega|)} & : \text{mag.}, \end{cases} \qquad (59)$$

where $\text{Re}\,G_d(\omega \to 0) \sim [\ln(16t/|\omega|)]^{-2}$ is used in the case of the magnetic impurity [obtained by Hilbert transform of Eq. (43)]. The divergence is thus sharper along the M-Γ line for magnetic impurities, as evident in Fig. 6.

Away from this divergent line (e.g., along the cut $\Gamma \to X \to M$ in Fig. 6), the QPI is characterized by vanishing scattering intensity at low scanning energies due to

$$\Lambda(q_x \neq q_y, \omega) \overset{|\omega|\to 0}{\sim} \tilde{a}_\mathbf{q} + \tilde{b}_\mathbf{q}|\omega| + i\tilde{c}_\mathbf{q}\omega \ln|\omega/t| + \dots, \qquad (60)$$

giving rise to the asymptotic behavior of the QPI,

$$|\Delta\rho(q_x \neq q_y, \omega)| \overset{|\omega|\to 0}{\sim} \begin{cases} \omega \ln|\omega/t| & : \text{scalar}, \\ \frac{1}{\ln(16t/|\omega|)} & : \text{mag.} \end{cases} \qquad (61)$$
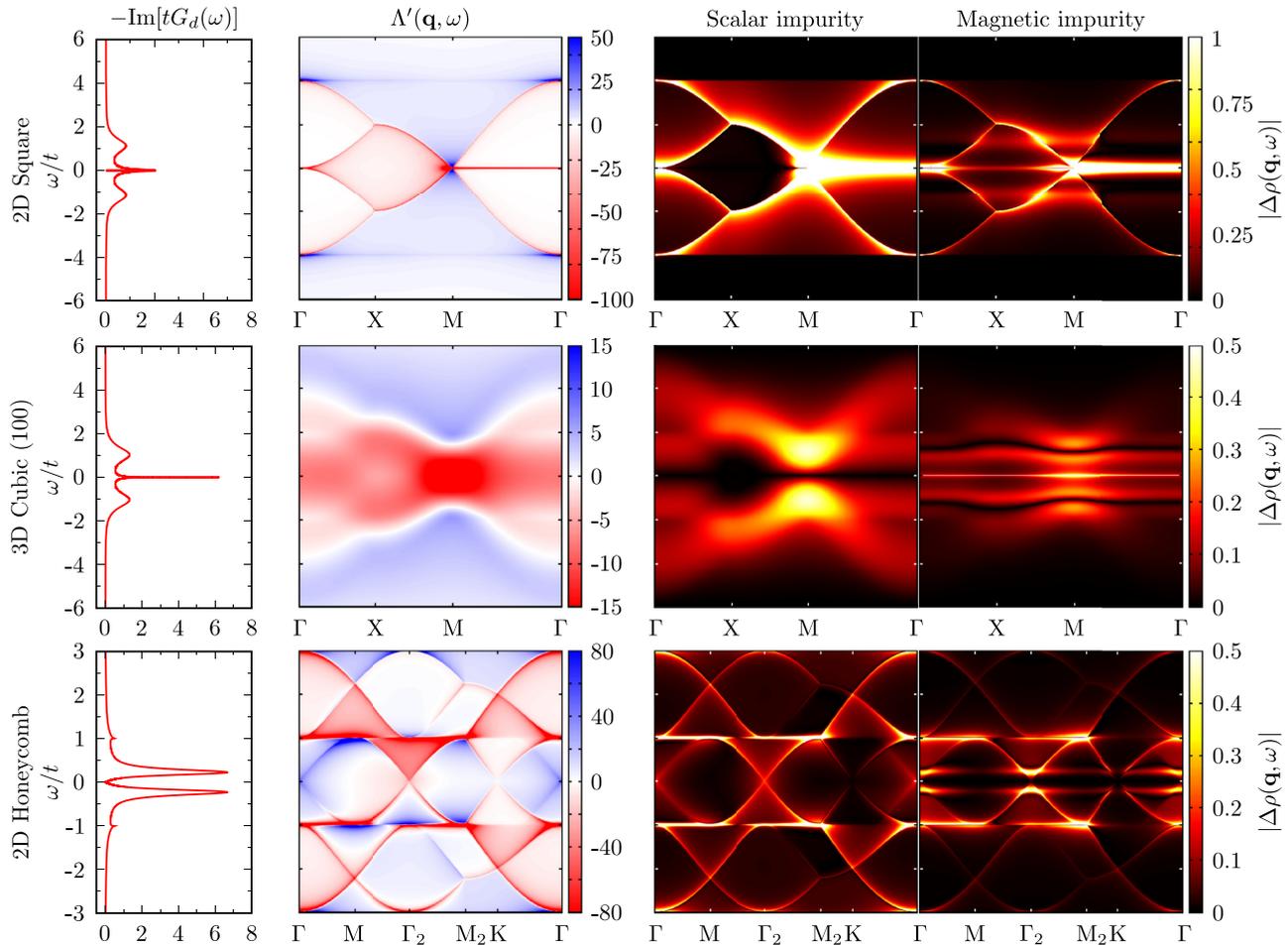
FIG. 6. (Color online) QPI dynamics for a single impurity on various lattices: 2D square, 3D cubic with (100) surface, honeycomb (top, middle, and bottom row panels, respectively). Right column panels: color plots of the QPI $|\Delta\rho(\mathbf{q},\omega)|$ across a Brillouin zone cut as a function of scanning energy ω/t, comparing scalar and magnetic impurities. Center column panels show a color plot of the host function $\Lambda'(\mathbf{q},\omega)$ over the same $\mathbf{q}$ cut and energies. Left column panels show the spectral function for a magnetic impurity $-\text{Im}[tG_d(\omega)]$ vs ω/t, calculated via NRG at $T = 0$. Magnetic impurity parameters as in Fig. 2. Scalar impurity $v = 0.15t$.

As a result, the QPI intensity for the magnetic impurity decays much more slowly than that of the scalar impurity as the scanning energy is reduced in the vicinity of the X symmetry point (see Fig. 6).

At higher scanning energies $|\omega| \sim U/2$, the "Hubbard satellites" in the spectral function of the magnetic impurity give rise to enhanced scattering near the M and X points. These features are not of course present in the scalar impurity QPI, and are as such one signature of strong electron correlations in magnetic impurities.

For the 3D cubic lattice, the dynamics are rather different, for two reasons: the host function $\Lambda(\mathbf{q},\omega)$ does not contain divergences, and the magnetic impurity Green function does not vanish at low energies because the host density of states is essentially flat for $|\omega| \ll t$.

For a scalar impurity on the cubic (100) surface, the QPI intensity vanishes everywhere at low energies $|\omega| \to 0$ because $\Lambda''(\mathbf{q},\omega) = -\Lambda''(\mathbf{q},-\omega)$ is odd in ω (see center row panels of Fig. 6). By contrast, the Kondo effect gives rise to enhanced scattering at temperatures and energies $\ll T_K$; this gives rise to a large finite QPI intensity for $\omega \ll T_K$. This is the

typical behavior expected for magnetic impurities in standard metallic systems.

Finally, we consider QPI dynamics on the the honeycomb lattice. In the case of the scalar impurity, the region of intense intravalley scattering around Γ and Γ_2 described by Eq. (55) disperses linearly at low energies according to Eq. (54), with $d_\Gamma(\omega) \sim \omega$. As $\omega \to 0$, the only divergent points are at $\mathbf{q}^* = \mathbf{q}_\Gamma$ and $\mathbf{q}_{\Gamma_2}$. At $\mathbf{q} = \mathbf{q}_K$, the QPI $\Delta\rho(\mathbf{q}_K,\omega) = 0$ is identically zero for any $|\omega| < t$. These features are shown in the lower panels of Fig. 6.

The QPI for the magnetic impurity near these points shows intense scattering at energies $|\omega| \sim U/2$, corresponding to the Hubbard satellites in the impurity spectral function due to charge fluctuations. At lower energies, however, $G_d(\omega)$ vanishes linearly according to Eq. (44) due to the host LDOS which also vanishes linearly at low energies. Importantly, the Kondo effect is suppressed at particle-hole symmetry, and the local moment phase is always stable for any interaction strength [24,28]. Electron correlations give rise to the nontrivial spin-flip scattering typical of such degenerate non-Fermi liquid phases [24,30]. As a result, the QPI for the magnetic

impurity vanishes everywhere at low energies, according to

$$|\Delta\rho(\mathbf{q},\omega)| \overset{|\omega|\to 0}{\sim} \begin{cases} |\omega|\ln^3|\omega/t| & : \mathbf{q} = \mathbf{q}_{\Gamma_{(2)}}, \\ 0 & : \mathbf{q} = \mathbf{q}_{\mathrm{K}}, \\ |\omega|\ln|\omega/t| & : \text{elsewhere}. \end{cases} \quad (62)$$

B. Universality

We focus now on a single magnetic impurity on the (100) surface of a 3D cubic lattice, the case most relevant to standard metallic systems where the host density of states becomes essentially flat at low energies [see Eq. (41)]. The Kondo effect is operative in such systems [1], with a spectral resonance setting in on temperature/energy scales $\sim T_K$ (see center panel of Fig. 2). This resonance embodies enhanced spin-flip scattering, which screens the impurity local moment dynamically. Importantly, all physical properties depend only on the single emergent scale T_K at low temperatures/energies, reflecting the *universal* RG flow between local moment and strong coupling fixed points [1].

At sufficiently low temperatures $T \ll T_K$, the impurity Green function $G_d(\omega)$ is a universal function of ω/T_K on all energy scales $|\omega| \ll \min(t, V^2/U)$, not only for $|\omega| \ll T_K$ where strong coupling Fermi liquid behavior Eq. (42) holds, but also for $|\omega| \gg T_K$ where local-moment physics dominates the dynamics. In that case, the impurity spectral function takes the asymptotic form [62]

$$-\operatorname{Im}G_d(\omega) \overset{|\omega|\gg T_K}{\sim} \frac{1}{1 + a\ln^2|b\omega/T_K|}, \quad (63)$$

with $a,b = \mathcal{O}(1)$ constants. This behavior for $T_K \ll |\omega| \ll \min(t, V^2/U)$ is universal because the hybridization function $\operatorname{Im}\Gamma(\omega)$ is essentially constant for $|\omega| \ll \min(t, V^2/U)$ on the cubic lattice [and $\operatorname{Re}\Gamma(\omega) \sim \omega$].

Similarly, the real part of the host function $\Lambda'(\mathbf{q},\omega)$ becomes constant on energy scales $|\omega| \ll t$, while $\Lambda'(\mathbf{q},\omega) \sim \omega$. In the scaling limit $T_K \to 0$, $\operatorname{Im}G_d(\omega)$ thus controls the energy dependence of the QPI in the universal regime [see Eq. (40)]. In practice, nonuniversal contributions are negligible for finite $T_K \ll \min(t, V^2/U)$.

In consequence, the entire QPI develops a universal scanning-energy dependence at low energies and temperatures. This means that magnetic impurities with different interaction strengths U and couplings V on different metallic substrates give the same normalized low-temperature/energy QPI $|\Delta\rho(\mathbf{q},\omega)|/|\Delta\rho(\mathbf{q},0)|$ when plotted versus rescaled scanning energy ω/T_K, for any scattering vector $\mathbf{q}$. Rescaled experimental QPI data for different systems should thus collapse onto a part of this universal curve, providing the unambiguous signature of scattering from magnetic impurities. This scaling collapse is demonstrated for the 3D cubic lattice in Fig. 7. Departure from universality is governed by the onset of ω dependence in $\Lambda(\mathbf{q},\omega)$, which is distinct for each $\mathbf{q}$; for the experimentally relevant parameters used, this is found to occur at $|\omega| \gtrsim 20T_K \simeq 0.1t$.

C. Thermal effects

So far, we have considered $T = 0$, appropriate in practice when $T \lesssim T_K \ll t$, such that temperature is the smallest energy scale in the problem. This regime is relevant, as typical
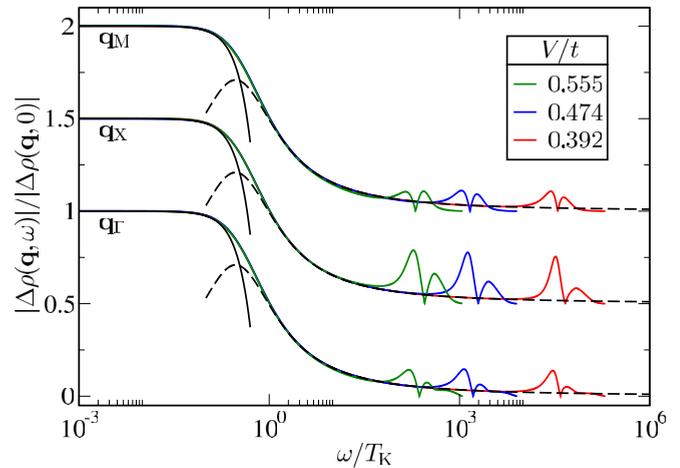
FIG. 7. (Color online) QPI $|\Delta\rho(\mathbf{q},\omega)|/|\Delta\rho(\mathbf{q},0)|$ vs ω/T_K for a magnetic impurity on the 3D cubic (100) surface for scattering vectors $\mathbf{q} = \mathbf{q}_\Gamma$, $\mathbf{q}_X$, and $\mathbf{q}_M$ (vertically offset by 0.5 for clarity), plotted for a range of impurity parameters: $V/t = 0.555, 0.474, 0.392$ with fixed $U = 1.95t$, corresponding to $T_K/t = 5\times10^{-3}$, 7×10^{-4}, and 3×10^{-5}. Equation (42) (solid line) and Eq. (63) (dashed line) describe $|\omega| \ll T_K$ and $\gg T_K$ asymptotes.

STM experiments are conducted at $\sim$5 K [14], but studies at higher T may also be performed [63] to investigate the change in QPI upon increasing temperature through T_K, and beyond.

Noninteracting conduction electrons and uncorrelated impurities (e.g., the scalar impurity) in practice have T-independent electronic structure, and thus QPI. By contrast, electronic correlations of the magnetic adatom exhibit a strong T dependence, entering the t matrix via the impurity Green function. As T is increased in metallic systems, the Kondo singlet is broken, destroying the Kondo resonance on the scale $T \gtrsim T_K$ and resulting in local-moment physics [1]. This results in a quite dramatic change in the t matrix, and hence QPI, on increasing T through T_K.

In addition to this interaction-driven T dependence, the local tunneling current measured in STS is weakly T dependent due to thermal excitation of conduction electrons; at finite T, the differential conductance

$$\frac{dI}{dV}(\mathbf{r}_i, \omega = eV, T) \propto \int_{-\infty}^{\infty} d\varepsilon\, \rho(\mathbf{r}_i, \varepsilon, T) f'(\omega - \varepsilon, T), \quad (64)$$

where $f'(\omega - \varepsilon, T) = \frac{d}{d\omega}f(\omega - \varepsilon, T)$ and $f(x, T) = [1 + \exp(x/T)]^{-1}$ is the Fermi function. Equation (64) represents the convolution of the LDOS (T dependent only for the magnetic impurity) with a broadening kernel, controlled by T. The QPI measured via FT-STS, $\Delta\rho_{\text{meas}}(\mathbf{q},\omega,T)$, is then related to the "pure" QPI by

$$\Delta\rho_{\text{meas}}(\mathbf{q},\omega,T) = \int_{-\infty}^{\infty} d\varepsilon\, \Delta\rho(\mathbf{q},\varepsilon,T) f'(\omega - \varepsilon, T). \quad (65)$$

Figure 8 shows the thermal evolution of the QPI for an impurity embedded on the 3D cubic (100) surface. For the magnetic impurity, the magnitude of the QPI decreases substantially as T increases through T_K, and the Kondo resonance is suppressed. By contrast, the QPI for a scalar impurity has a much simpler T dependence [entering only
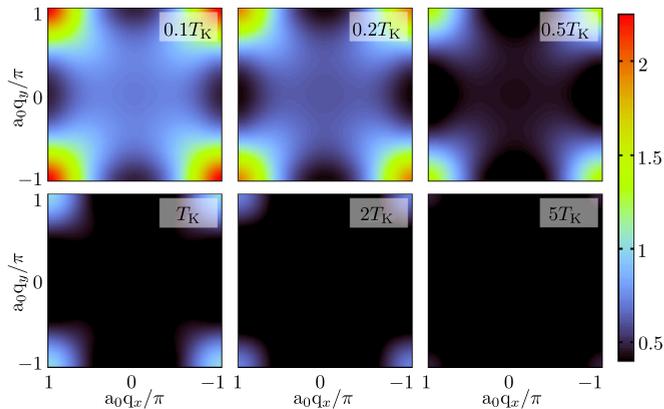
FIG. 8. (Color online) Simulated experimental FT-STS measurements $|\Delta\rho_{\text{meas}}(\mathbf{q}, \omega = 0.1T_K)|$ for a single magnetic impurity (parameters as in Fig. 2, such that $T_K = 5 \times 10^{-3}t$) at fixed scanning energy, plotted across the 1BZ for a series of temperatures $T/T_K = 0.1, 0.2, 0.5, 1, 2,$ and 5.

via the thermal broadening, Eq. (65)], with essentially no T dependence for $T \ll t$, as depicted by the dashed line in Fig. 9. The strong T dependence of QPI is a characteristic signature of Kondo physics in systems with magnetic impurities.

The asymptotic low-T expansion of the impurity Green function at particle-hole symmetry and $\omega = 0$ (applicable in the regime $\omega \ll T \ll T_K$) is a universal function of the Kondo temperature [1]

$$\text{Im}\Gamma(\omega = 0) \times \text{Im}\, G_d(\omega = 0, T) \overset{T \ll T_K}{\sim} 1 - \alpha_T(T/T_K)^2 + \cdots.$$

(66)
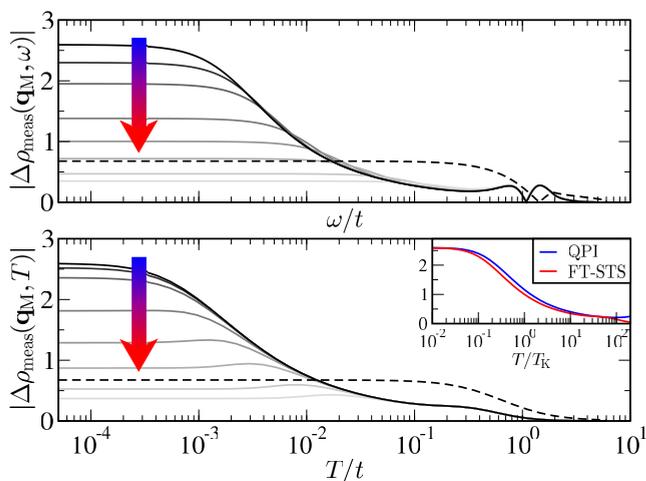
FIG. 9. (Color online) QPI $|\Delta\rho_{\text{meas}}(\mathbf{q}_M)|$ for a single magnetic impurity (parameters given in Fig. 2, such that $T_K = 5 \times 10^{-3}t$) at the point $\mathbf{q}_M = (\pi, \pi)$ in the 1BZ, shown as a function of ω (T) at a series of different T (ω) in the upper (lower) panel, such that $T/T_K(\omega/T_K) = 0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10$. Increasing T from low to high is denoted by the arrow. The QPI due to a scalar impurity is also plotted (dashed line) for comparison. Lower panel inset: the measured FT-STS (red) and the "pure" QPI (blue) vs T/T_K, showing the same universal behavior in each case (up to a scale factor due to thermal broadening).

This leads to a close correspondence between the ω and T dependencies of the QPI [via the impurity Green function, comparing Eqs. (42) and (66)], as seen by comparison of the upper and lower panels in Fig. 9.

As such, we expect to observe universal scaling in the T dependence of the QPI for the magnetic impurity, analogous to that of the ω dependence. The comparison of the pure QPI and the thermally broadened FT-STS signal in the inset of Fig. 9 demonstrates this universal behavior, which is unaffected by the thermal broadening of the STM-measured conductance Eq. (65) (up to a trivial scale factor).

VI. INTERPRETATION OF FT-STS

A. Finite-size effects

In experiment, the surface LDOS $\Delta\rho(\mathbf{r}_i, \omega)$ is measured over an $L \times L$ plaquette using STM [40], with the QPI obtained from Eq. (15). The $\mathbf{q}$-space resolution of the resulting QPI naturally depends on the real-space sample size. The "true" QPI is recovered as $L \to \infty$, obtained theoretically by the t-matrix approach.

We now consider explicitly the effects of finite sample size, by simulating the experimental protocol. The LDOS for these surface sites is calculated exactly using Eq. (16), with the nonlocal free Green functions obtained using the convolution method described in the Appendix.

Figure 10 shows a Brillouin zone cut through the QPI for a magnetic impurity on the cubic lattice (100) surface, computed in the $L \to \infty$ limit using the t-matrix approach (solid line). This true QPI is compared with results for the same system restricting to an $L \times L$ surface sample, with $L = 100$ (crosses) and $L = 10$ (diamonds). The true QPI is very well approximated when $L = 100$ is used (corresponding to a plaquette of side length $\sim 10^2$ Å, as typical in experiment [40]). The $\mathbf{q}$-space resolution is also sufficient to capture accurately all features. Indeed, even for an extremely small sample region $L = 10$, the accuracy is surprisingly good, although the discretization is severe.

Reassuringly, the experimental protocol reproduces accurately the true QPI. A large sample size is, however, still
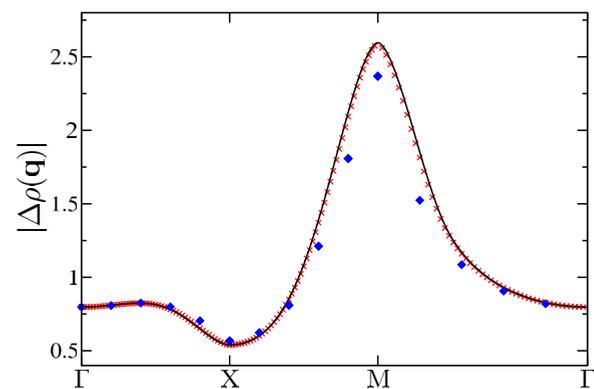
FIG. 10. (Color online) Brillouin zone cut of $|\Delta\rho(\mathbf{q}, \omega)|$ for a magnetic impurity on the 3D cubic (100) surface at $\omega = 10^{-4}t$. Exact QPI (line) calculated via the t-matrix approach [Eq. (40)], compared with the direct real-space approach [Eq. (15)], sampling an $L \times L$ surface plaquette with $L = 100$ (cross points) and $L = 10$ (diamond points). Impurity parameters as in Fig. 2.

needed to resolve sharp **q**-space features; in 2D systems, the characteristic sharp cusps in the QPI in Figs. 3 and 5 would require very large LDOS samples in real space.

B. JDOS interpretation

The QPI $\Delta\rho(\mathbf{q},\omega)$ is often interpreted (e.g., Refs. [13,40]) in terms of the joint density of states [JDOS, $J^I(\mathbf{q},\omega)$], viz. $\Delta\rho(\mathbf{q},\omega) \simeq J^I(\mathbf{q},\omega)$ with

$$J^I(\mathbf{q},\omega) = \int_{1\text{SBZ}} \frac{d^2\mathbf{k}}{\Omega_{\text{BZ}}} \rho^0(\mathbf{k},\omega)\rho^0(\mathbf{k}-\mathbf{q},\omega) \qquad (67)$$

and $\rho^0(\mathbf{k},\omega) = -\frac{1}{\pi}\text{Im}G^0(\mathbf{k},\omega)$ the **k**-resolved surface density of states at energy ω. Contributions to the JDOS arise when quasiparticles on a constant energy contour with momenta **k** and **k**′ are separated by $\mathbf{q} = \mathbf{k} - \mathbf{k}'$.

On heuristic grounds, it is usually argued that the amplitude of impurity-induced scattering from **k** to **k**′ at energy ω can only be significant if there is a high density of quasiparticle states at both **k** *and* **k**′ (i.e., they have a large JDOS). The QPI at scattering vector **q** is the sum of all scattering processes where $\mathbf{k}' = \mathbf{k} - \mathbf{q}$. By assuming that the QPI is large when the JDOS is large, QPI patterns can be used to infer the JDOS and hence electronic properties of the clean host material. Experimental QPI data are typically interpreted in this way [13,40].

The usefulness and relative simplicity of the JDOS picture has motivated efforts to connect rigorously the QPI and JDOS. Notably, the perturbative approach employed in Ref. [64] makes the link by assuming a constant scattering amplitude and phase along the constant energy contour. But, a faithful description of interfering scattering processes typically requires relative phase information, and the JDOS simply lacks information about overlap matrix elements between states in the impurity-coupled system. In consequence, the QPI may be small even when the JDOS is large (as may be verified explicitly). The JDOS picture then fails to capture the basic physics of the scattering, as is known, e.g., in graphene [13]. We emphasize that $J^I(\mathbf{q},\omega)$ cannot be derived from $\Delta\rho(\mathbf{q},\omega)$ at a given **q** and ω, in any controlled limit.

However, for a simple scalar impurity on a centrosymmetric lattice, the QPI and JDOS are in fact related by Hilbert transformation. From Eqs. (38) and (45) [or, e.g., Eq. (49)], the QPI in this case can be written as

$$\Delta\rho(\mathbf{q},\omega) = 2v \int_{1\text{SBZ}} \frac{d^2\mathbf{k}}{\Omega_{\text{BZ}}} [\text{Re}G^0(\mathbf{k},\omega)]\rho^0(\mathbf{k}-\mathbf{q},\omega), \qquad (68)$$

where we have exploited periodicity across the 1SBZ. We introduce the complex quantity $J(\mathbf{q},\omega)$, defined as

$$J(\mathbf{q},\omega) = -\frac{1}{\pi} \int_{1\text{SBZ}} \frac{d^2\mathbf{k}}{\Omega_{\text{BZ}}} G^0(\mathbf{k},\omega)\rho^0(\mathbf{k}-\mathbf{q},\omega) \qquad (69a)$$

$$\equiv -\frac{1}{2v\pi}\Delta\rho(\mathbf{q},\omega) + i J^I(\mathbf{q},\omega), \qquad (69b)$$

such that the QPI and JDOS are Hilbert conjugates, being respectively the real and imaginary parts of $J(\mathbf{q},\omega)$.

The JDOS interpretation of the QPI may thus be roughly correct for dilute scalar impurities on centrosymmetric lattices because the underlying **q**-space topology of singular and nodal lines is the same for $J^I(\mathbf{q},\omega)$ as it is for $\Delta\rho(\mathbf{q},\omega)$, both
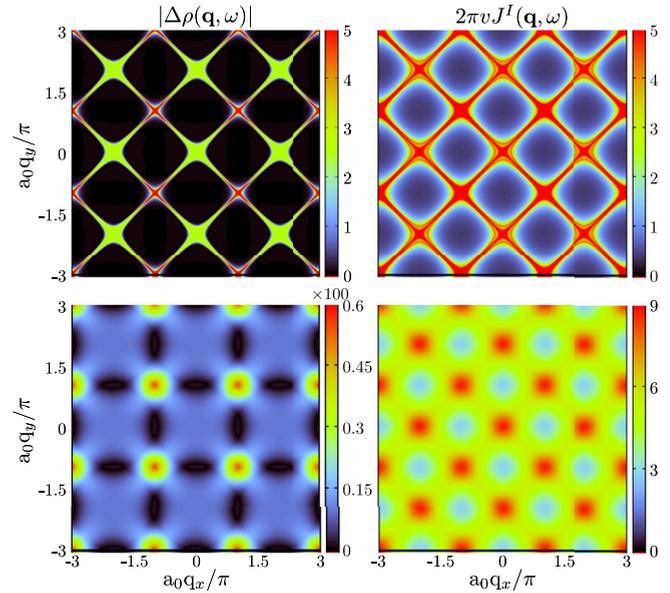
FIG. 11. (Color online) Comparison of QPI, $|\Delta\rho(\mathbf{q},\omega)|$ (left column), with the JDOS (right column) plotted as $2\pi v J^I(\mathbf{q},\omega)$ [see Eq. (69)], for the 2D square lattice (upper row) and (100) surface of 3D cubic lattice (lower row), for scanning energy ω and Born impurity scattering potential v as in Figs. 3 and 4, respectively.

being related to a single complex analytic function $J(\mathbf{q},\omega)$. Nevertheless, even in this case the QPI is *not* accessible directly from the JDOS at a given **q** and ω: the *entire* ω dependence of $J^I(\mathbf{q},\omega)$ must be known to obtain $\Delta\rho(\mathbf{q},\omega)$ by Hilbert transformation. By way of illustration, Fig. 11 compares the JDOS to the QPI for the 2D square and (100) surface of the 3D cubic lattice; in the former case, the two quantities are roughly similar, while in the latter the JDOS has significantly different **q**-space structure and intensity.

We add that the above connection [Eq. (69b)] does *not* hold in the case of noncentrosymmetric lattices or bipartite lattices, where Eq. (68) is inapplicable. The QPI, $\Delta\rho(\mathbf{q},\omega)$, then depends on both real and imaginary parts of $J(\mathbf{q},\omega)$ due to additional **q**-dependent phase factors. This explains the failure of the JDOS interpretation in the case of impurities in graphene [13] (see Sec. IV D).

Finally, we emphasize that the JDOS and QPI are never related simply by Hilbert transformation when magnetic impurities are present. This is because the t matrix is a complex dynamical object: the real and imaginary parts of $J(\mathbf{q},\omega)$ are again mixed. Indeed, the Hilbert transform, involving integration over all energy scales, necessitates a full knowledge of the impurity dynamics, information simply not contained in the JDOS.

VII. CONCLUSION

We have studied theoretically the use of quasiparticle interference (QPI), measured in FT-STS experiments, as a probe of magnetic adatoms on surfaces. Following a general formulation of the QPI due to an arbitrary distribution of impurities, we turned explicitly to single impurities adsorbed on a range of host surfaces: the (100) surface of a 3D simple cubic lattice, and the 2D honeycomb and square lattices, in which the

Fermi level densities of states, respectively, embody standard metallic behavior, pseudogap behavior, and a divergence due to a Van Hove singularity. In all cases, the single-impurity QPI factorizes into a local scattering t matrix, and a host response function $\Lambda(\mathbf{q},\omega)$ at scanning energy $\omega = eV$.

The scattering t matrix for a magnetic impurity is simply related to the impurity Green function, itself dependent on the host lattice, and thus the rich dynamics due to electronic correlations is manifest in the QPI. Despite the local, momentum-independent nature of these correlations, the $\mathbf{q}$-space structure of the QPI is found to be qualitatively different from that of a simple scalar impurity due to nontrivial phase shifts associated with scattering from magnetic impurities, which reflect, e.g., the Kondo effect.

The response function $\Lambda(\mathbf{q},\omega)$ is also nontrivial, despite being a property of the free, noninteracting host. It displays significant structure in $\mathbf{q}$ space, symptomatic of the symmetry and dimensionality of the host, but its energy dependence becomes featureless for $|\omega| \ll t$ (with t the intersite lattice hopping). By contrast, the Kondo physics due to a magnetic impurity is controlled by an emergent scale $T_K \ll t$, and so scattering becomes strongly energy dependent at low energies. Indeed, the QPI exhibits *universal* scaling in terms of ω/T_K and T/T_K, a characteristic hallmark for systems containing magnetic impurities. Conversely, the QPI for systems containing scalar impurities has no energy or temperature dependence on scales $\ll t$.

The more complex case of QPI for multiple, mutually interacting magnetic impurities remains to be investigated. Such systems will display an even wider array of impurity physics, due to the competition of local and nonlocal (RKKY-type) interactions between impurities. These are expected to have a significant impact on the QPI for randomly distributed impurities, and is the subject of future work [23].

APPENDIX: CONVOLUTION METHOD FOR LATTICE GREEN FUNCTIONS

The calculation of real-space lattice Green functions (LGFs) for periodic tight-binding (TB) models is a well-known problem relevant to many areas of physics [46,48,49]. The (retarded) real-space Green function between sites at $\mathbf{r}$ and $\mathbf{r}'$ on a d-dimensional lattice is generally given by

$$G^0(\mathbf{r},\mathbf{r}',\omega) = \int_{1BZ} \frac{d^d\mathbf{k}}{\Omega_{BZ}} \frac{e^{-i(\mathbf{r}'-\mathbf{r}).\mathbf{k}}}{\omega + i0^+ - \epsilon_0 - \epsilon_\mathbf{k}}, \quad (A1)$$

where $\epsilon_\mathbf{k}$ is the dispersion and $\epsilon_0 \equiv \mu$ is a constant onsite energy (or chemical potential). For a nearest-neighbor (NN) TB lattice specified by Eq. (1), $\epsilon_\mathbf{k} = t \sum_n e^{i\delta_n.\mathbf{k}}$, with $\{\delta_n\}$ the set of NN lattice vectors.

In 1D, simple expressions for the local (onsite) and nonlocal (intersite) LGFs are readily obtained, either by direct evaluation of Eq. (A1) or via equations of motion [38,65]. For the terminal site of a semi-infinite 1D chain with onsite energies ϵ_0, the local LGF is given exactly in closed form by

$$G^0_{1D}(\epsilon_0,\omega) = f\left(\frac{\omega - \epsilon_0}{2t}\right) \quad \text{where}$$

$$tf(\tilde\omega) = \tilde\omega - \begin{cases} \text{sgn}(\tilde\omega)\sqrt{\tilde\omega^2 - 1}, & |\tilde\omega| > 1 \\ i\sqrt{1 - \tilde\omega^2}, & |\tilde\omega| \leqslant 1 \end{cases} \quad (A2)$$

which is equivalent to Eq. (25). The LGFs for the *infinite* 1D chain can be obtained in terms of Eq. (A2) by exploiting translational invariance

$$G^0_{1D_\infty}(x,x',\epsilon_0,\omega) = \frac{\left[tG^0_{1D}(\epsilon_0,\omega)\right]^{|x-x'|}}{\omega - \epsilon_0 - 2t^2 G^0_{1D}(\epsilon_0,\omega)}. \quad (A3)$$

However, LGFs for various lattice geometries in two and three dimensions are typically highly complicated [47,50] and not available in closed form. Direct numerical evaluation of Eq. (A1) is notoriously difficult, particularly for low energies, large site separations, or in the vicinity of Van Hove singularities. Recursion relations have been established in several cases, but solutions are often numerically unstable [51]. Improved variants of the recursion technique (or continued fraction expansions) have been developed [51,53], but are costly to implement if LGFs are needed as an entire function of frequency.

Here, we derive an approach to the calculation of LGFs on hypercubic-type lattices, which is both highly accurate and numerically efficient. The method exploits the simple closed-form expressions for the LGFs in 1D [Eqs. (A2) and (A3)], building up lattices in higher dimensions by successive convolutions of those functions. The process is highly efficient because fast Fourier transform algorithms can be used to perform the convolution integrals. The method also has the advantage that boundary edges in 2D systems or explicit surfaces in 3D systems can be simply treated. Indeed, 2D nanoribbon or 3D block geometries can be implemented, and infinite systems can also be handled directly with no extra cost.

We demonstrate the method first for the infinite 2D square lattice (and henceforth set $\epsilon_0 = 0$ without loss of generality). We denote the creation operator at lattice site $\mathbf{r} = (x,y)$ as $c^\dagger_{\mathbf{r},\sigma} \equiv c^{x\,\dagger}_{y,\sigma}$, and define the vector of operators for row y as $\vec{c}^{\,\dagger}_{y,\sigma} = (\ldots, c^{1\,\dagger}_{y,\sigma}, c^{2\,\dagger}_{y,\sigma}, c^{3\,\dagger}_{y,\sigma}, \ldots)$. The TB Hamiltonian then takes the form

$$\mathcal{H}_{2D} = \sum_{y=-\infty}^{\infty} \sum_\sigma [\vec{c}^{\,\dagger}_{y,\sigma} \hat{T}_{1D} \vec{c}_{y,\sigma} - t(\vec{c}^{\,\dagger}_{y,\sigma} \vec{c}_{y+1,\sigma} + \text{H.c.})], \quad (A4)$$

where $\hat{T}_{1D}$ a matrix describing the connectivity between sites of the yth row (here equal to t for nearest-neighbor sites and 0 otherwise). Importantly, for hypercubic-type lattices, $\hat{T}_{1D}$ is independent of row index y. Equation (A4) represents a set of *coupled* infinite 1D TB chains to form the 2D lattice.

We now perform a canonical transformation of operators $\vec{f}_{y,\sigma} = \hat{U}^\dagger \vec{c}_{y,\sigma}$, with the matrix $\hat{U}$ defined such that $\hat{D} = \hat{U}^\dagger \hat{T}_{1D} \hat{U}$ is diagonal ($D_{kk'} = \epsilon^{1D}_k \delta_{kk'}$). Since this system is

(a) 2D square lattice

(b) 3D cubic lattice with (100) surface
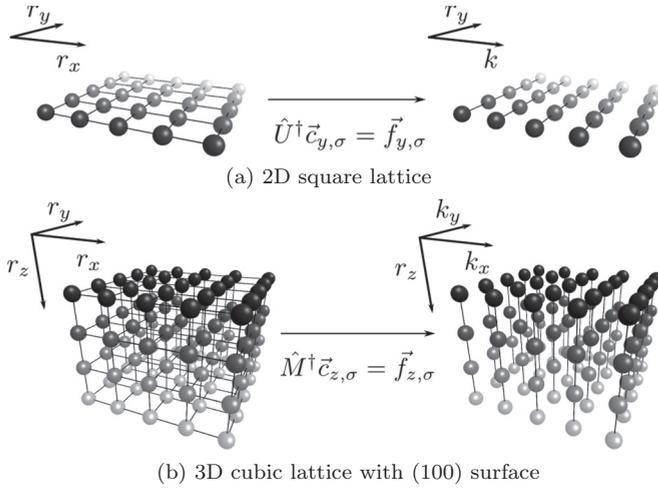
FIG. 12. Schematic showing the transformation from the real-space basis (left) to a basis of *decoupled* 1D chains (right). Bonds denote hoppings t connecting sites. $\hat{U}$ diagonalizes rows of constant r_y in the infinite 2D square lattice system of (a), while $\hat{M}$ diagonalizes planes of constant $r_z \geq 0$ in the semi-infinite 3D cubic lattice system of (b).

infinite and periodic in the x direction, $k \equiv k_x$ can be understood as the Bloch momentum, and ϵ_k^{1D} the 1D dispersion. However, in general (e.g., for systems with a boundary), k merely labels an eigenstate of $\hat{U}$, with eigenvalue ϵ_k^{1D}.

In this basis, Eq. (A4) reduces to

$$\mathcal{H}_{2D} = \sum_k \left[\sum_{y,\sigma} \epsilon_k^{1D} f_{y,\sigma}^{k\,\dagger} f_{y,\sigma}^k - t\left(f_{y,\sigma}^{k\,\dagger} f_{y+1,\sigma}^k + \text{H.c.} \right) \right], \quad (A5)$$

which describes a set of *decoupled* 1D chains labeled by k, each with constant onsite energy ϵ_k^{1D}. The transformation from coupled to uncoupled chains is shown schematically in Fig. 12(a).

The LGFs can then be expressed as

$$\begin{aligned}
G_{2D}^0(\mathbf{r},\mathbf{r}',\omega) &\equiv \left\langle\!\left\langle c_{y,\sigma}^x ; c_{y',\sigma}^{x'\,\dagger} \right\rangle\!\right\rangle_\omega \\
&= \sum_{k,k'} U_{xk} U_{k'x'}^* \left\langle\!\left\langle f_{y,\sigma}^k ; f_{y',\sigma}^{k'\,\dagger} \right\rangle\!\right\rangle_\omega \\
&= \sum_k U_{xk} U_{kx'}^* G_{1D_\infty}^0\left(y,y',\epsilon_k^{1D},\omega\right), \quad (A6)
\end{aligned}$$

where the last line follows because $\left\langle\!\left\langle f_{y,\sigma}^k ; f_{y',\sigma}^{k'\,\dagger} \right\rangle\!\right\rangle_\omega \propto \delta_{kk'}$ is diagonal in k [see Fig. 12(a)]. The 1D Green function $G_{1D_\infty}^0(y,y',\epsilon_k^{1D},\omega)$ is given by Eq. (A3).

We now make use of the spectral representation of the 1D LFGs, which can be expressed in terms of ϵ_k^{1D} and U_{xk}, viz.,

$$\text{Im}\,G_{1D_\infty}^0(x,x',0,\omega') = -\pi \sum_k U_{xk} U_{kx'}^* \delta\left(\omega' - \epsilon_k^{1D}\right). \quad (A7)$$

One can then write Eq. (A6) as

$$\begin{aligned}
G_{2D}^0(\mathbf{r},\mathbf{r}',\omega) = &-\frac{1}{\pi} \int_{-\infty}^\infty d\omega' \text{Im}\,G_{1D_\infty}^0(x,x',0,\omega') \\
&\times G_{1D_\infty}^0(y,y',\omega',\omega). \quad (A8)
\end{aligned}$$

Since $G_{1D_\infty}^0(y,y',\omega',\omega) \equiv G_{1D_\infty}^0(y-y',\omega-\omega')$ from Eqs. (A2) and (A3), Eq. (A8) takes the form of a convolution integral. Convolution theorem then allows the efficiency of fast Fourier transform algorithms to be exploited, as

$$\begin{aligned}
\mathcal{F}_\omega[G^0(\mathbf{r},\mathbf{r}',\omega)] = &-\frac{1}{\pi} \mathcal{F}_\omega\left[\text{Im}\,G_{1D_\infty}^0(x,x',\omega)\right] \\
&\times \mathcal{F}_\omega\left[G_{1D_\infty}^0(y,y',\omega)\right], \quad (A9)
\end{aligned}$$

where $\mathcal{F}_\omega$ denotes Fourier transformation.

Straightforward extension of this method allows access to LGFs in higher dimensions. As a final instructive example, we consider now the 3D cubic lattice with an explicit (100) surface. The Hamiltonian is written as

$$\mathcal{H}_{3D} = \sum_{z=0}^\infty \sum_\sigma [\vec{c}_{z,\sigma}^{\,\dagger} \hat{T}_{2D} \vec{c}_{z,\sigma} - t(\vec{c}_{z,\sigma}^{\,\dagger} \vec{c}_{z+1,\sigma} + \text{H.c.})], \quad (A10)$$

in terms of vectors of operators for *planes* stacked in the z direction, $\vec{c}_{z,\sigma}^{\,\dagger} = (\ldots, c_{z,\sigma}^{\mathbf{r}_1\,\dagger}, c_{z,\sigma}^{\mathbf{r}_2\,\dagger}, c_{z,\sigma}^{\mathbf{r}_3\,\dagger}, \ldots)$, where $c_{z,\sigma}^{\mathbf{r}\,\dagger}$ creates an electron at site $\mathbf{r} = (x,y)$ of plane z. $\hat{T}_{2D}$ is now the connectivity matrix for the 2D square lattice planes.

We now diagonalize each 2D plane by writing $\vec{f}_{z,\sigma} = \hat{M}^\dagger \vec{c}_{z,\sigma}$ such that $\hat{D} = \hat{M}^\dagger \hat{T}_{2D} \hat{M}$ is diagonal. As before, $D_{kk'} = \epsilon_k^{2D} \delta_{kk'}$, but now ϵ_k^{2D} is the 2D square lattice dispersion. In the transformed basis, the semi-infinite 3D cubic lattice becomes a bundle of decoupled semi-infinite 1D chains, each with onsite energy ϵ_k^{2D}, as depicted in Fig. 12(b):

$$\mathcal{H}_{3D} = \sum_{k,\sigma} \left[\sum_{z=0}^\infty \epsilon_k^{2D} f_{z,\sigma}^{k\,\dagger} f_{z,\sigma}^k - t\left(f_{z,\sigma}^{k\,\dagger} f_{z+1,\sigma}^k + \text{H.c.} \right) \right]. \quad (A11)$$

The surface LGFs, with $z = 0$, then follow as

$$\begin{aligned}
G_{\text{surf}}^0(\mathbf{r},\mathbf{r}',\omega) &\equiv \left\langle\!\left\langle c_{0,\sigma}^{\mathbf{r}} ; c_{0,\sigma}^{\mathbf{r}'\,\dagger} \right\rangle\!\right\rangle_\omega \\
&= \sum_k M_{\mathbf{r}k} M_{k\mathbf{r}'}^* G_{1D}^0\left(\epsilon_k^{2D},\omega\right), \quad (A12)
\end{aligned}$$

where $G_{1D}^0(\epsilon_k^{2D},\omega)$ is given by Eq. (A2). Employing the spectral representation of the 2D square lattice Green functions,

$$\text{Im}\,G_{2D}^0(\mathbf{r},\mathbf{r}',\omega) = -\pi \sum_k M_{\mathbf{r}k} M_{k\mathbf{r}'}^* \delta\left(\omega - \epsilon_k^{2D}\right), \quad (A13)$$

we can write

$$G_{\text{surf}}^0(\mathbf{r},\mathbf{r}',\omega) = -\frac{1}{\pi} \int_{-\infty}^\infty d\omega' \text{Im}\,G_{2D}^0(\mathbf{r},\mathbf{r}',\omega') G_{1D}^0(\omega',\omega), \quad (A14)$$

which can again be viewed as a convolution, here between the semi-infinite 1D Green function $G_{1D}^0(\omega',\omega) \equiv G_{1D}^0(\omega-\omega')$ given in Eq. (A2) and the 2D square lattice Green function $G_{2D}^0(\mathbf{r},\mathbf{r}',\omega)$ given in Eq. (A8).

[1] A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge, 1997).

[2] J. Li, W.-D. Schneider, R. Berndt, and B. Delley, Phys. Rev. Lett. **80**, 2893 (1998); M. Ternes, A. J. Heinrich, and W.-D. Schneider, J. Phys.: Condens. Matter **21**, 053001 (2009).

[3] V. Madhavan, W. Chen, T. Jamneala, M. F. Crommie, and N. S. Wingreen, Science **280**, 567 (1998); Phys. Rev. B **64**, 165412 (2001).

[4] H. C. Manoharan, C. P. Lutz, and D. M. Eigler, Nature (London) **403**, 512 (2000).

[5] A. Heinrich and S. Loth, Science **332**, 1039 (2011).

[6] A. A. Khajetoorians, J. Wiebe, B. Chilian, and R. Wiesendanger, Science **332**, 1062 (2011).

[7] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).

[8] R. M. Feenstra, J. A. Stroscio, and A. P. Fein, Surf. Sci. **181**, 295 (1987).

[9] Q.-H. Wang and D.-H. Lee, Phys. Rev. B **67**, 020511 (2003).

[10] L. Capriotti, D. J. Scalapino, and R. D. Sedgewick, Phys. Rev. B **68**, 014508 (2003).

[11] R. S. Markiewicz, Phys. Rev. B **69**, 214517 (2004).

[12] K. McElroy, G.-H. Gweon, S. Y. Zhou, J. Graf, S. Uchida, H. Eisaki, T. Takagi, T. Sasagawa, D.-H. Lee, and A. Lanzara, Phys. Rev. Lett. **96**, 067005 (2006).

[13] L. Simon, C. Bena, F. Vonau, M. Cranney, and D. Aubel,, J. Phys. D: Appl. Phys. **44**, 464010 (2011).

[14] P. Wahl, L. Diekhöner, M. A. Schneider, L. Vitali, G. Wittich, and K. Kern, Phys. Rev. Lett. **93**, 176603 (2004).

[15] P. Anderson, Phys. Rev. **124**, 41 (1961).

[16] A. Schiller and S. Hershfield, Phys. Rev. B **61**, 9036 (2000).

[17] O. Újsághy, J. Kroha, L. Szunyogh, and A. Zawadowski, Phys. Rev. Lett. **85**, 2557 (2000).

[18] R. Zitko, Phys. Rev. B **84**, 195116 (2011).

[19] R. Bulla, T. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008).

[20] T. A. Costi, L. Bergqvist, A. Weichselbaum, J. von Delft, T. Micklitz, A. Rosch, P. Mavropoulos, P. H. Dederichs, F. Mallet, L. Saminadayar, and C. Bäuerle, Phys. Rev. Lett. **102**, 056802 (2009); M. Hanl, A. Weichselbaum, T. A. Costi, F. Mallet, L. Saminadayar, C. Bäuerle, and J. von Delft, Phys. Rev. B **88**, 075146 (2013).

[21] L. Huang, T. O. Wehling, and P. Werner, Phys. Rev. B **89**, 245104 (2014).

[22] S. Gardonio, M. Karolak, T. O. Wehling, L. Petaccia, S. Lizzit, A. Goldoni, A. I. Lichtenstein, and C. Carbone, Phys. Rev. Lett. **110**, 186404 (2013).

[23] A. K. Mitchell, P. Derry, and D. E. Logan, Phys. Rev. B **91**, 235127 (2015).

[24] C. Gonzalez-Buxton and K. Ingersent, Phys. Rev. B **57**, 14254 (1998).

[25] A. K. Mitchell, D. Schuricht, M. Vojta, and L. Fritz, Phys. Rev. B **87**, 075430 (2013).

[26] A. K. Mitchell, M. Vojta, R. Bulla, and L. Fritz, Phys. Rev. B **88**, 195119 (2013); A. K. Mitchell and L. Fritz, *ibid.* **88**, 075104 (2013).

[27] C. A. Busser, G. B. Martins, and A. E. Feiguin, Phys. Rev. B **88**, 245113 (2013).

[28] L. Fritz and M. Vojta, Rep. Prog. Phys. **76**, 032501 (2013).

[29] R. Bulla, M. T. Glossop, D. E. Logan, and T. Pruschke, J. Phys.: Condens. Matter **12**, 4899 (2000).

[30] D. E. Logan, A. P. Tucker, and M. R. Galpin, Phys. Rev. B **90**, 075150 (2014).

[31] A. W. W. Ludwig, M. P. A. Fisher, R. Shankar, and G. Grinstein, Phys. Rev. B **50**, 7526 (1994).

[32] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).

[33] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975); H. R. Krishna-murthy, J. W. Wilkins, and K. G. Wilson, Phys. Rev. B **21**, 1003 (1980); H. R. Krishna-murthy, J. W. Wilkins, and K. G. Wilson, *ibid.* **21**, 1044 (1980).

[34] R. Bulla, A. C. Hewson, and T. Pruschke, J. Phys.: Condens. Matter **10**, 8365 (1998).

[35] A. Weichselbaum and J. von Delft, Phys. Rev. Lett. **99**, 076402 (2007).

[36] R. Peters, T. Pruschke, and F. B. Anders, Phys. Rev. B **74**, 245114 (2006).

[37] F. B. Anders and A. Schiller, Phys. Rev. Lett. **95**, 196801 (2005).

[38] D. N. Zubarev, Sov. Phys. Usp **3**, 320 (1960).

[39] L. Petersen, P. T. Sprunger, P. Hofmann, E. Lægsgaard, B. G. Briner, M. Doering, H.-P. Rust, A. M. Bradshaw, F. Besenbacher, and E. W. Plummer, Phys. Rev. B **57**, R6858 (1998); L. Petersen, P. Hofmann, E. W. Plummer, and F. Besenbacher, J. Electron. Spectrosc. Relat. Phenom. **109**, 97 (2000).

[40] J. E. Hoffman, K. McElroy, D.-H. Lee, K. M. Lang, H. Eisaki, S. Uchida, and J. C. Davis, Science **297**, 1148 (2002).

[41] C. Howald, H. Eisaki, N. Kaneko, M. Greven, and A. Kapitulnik, Phys. Rev. B **67**, 014533 (2003); O. Fischer, M. Kugler, I. Maggio-Aprile, C. Berthod, and C. Renner, Rev. Mod. Phys. **79**, 353 (2007).

[42] Y. Yin, M. Zech, T. L. Williams, X. F. Wang, G. Wu, X. H. Chen, and J. E. Hoffman, Phys. Rev. Lett. **102**, 097002 (2009); Y. Yin, M. Zech, T. L. Williams, and J. E. Hoffman, Phys. C (Amsterdam) **469**, 535 (2009); J. T. Park, D. S. Inosov, C. Niedermayer, G. L. Sun, D. Haug, N. B. Christensen, R. Dinnebier, A. V. Boris, A. J. Drew, L. Schulz, T. Shapoval, U. Wolff, V. Neu, X. Yang, C. T. Lin, B. Keimer, and V. Hinkov, Phys. Rev. Lett. **102**, 117006 (2009).

[43] P. Roushan, J. Seo, C. V. Parker, Y. S. Hor, D. Hsieh, D. Qian, A. Richardella, M. Z. Hasan, R. J. Cava, and A. Yazdani, Nature (London) **460**, 1106 (2009); T. Zhang, P. Cheng, X. Chen, J.-F. Jia, X. Ma, K. He, L. Wang, H. Zhang, X. Dai, Z. Fang, X. Xie, and Q.-K. Xue, Phys. Rev. Lett. **103**, 266803 (2009).

[44] I. Brihuega, P. Mallet, C. Bena, S. Bose, C. Michaelis, L. Vitali, F. Varchon, L. Magaud, K. Kern, and J. Y. Veuillen, Phys. Rev. Lett. **101**, 206802 (2008); Y. Zhang, V. W. Brar, C. Girit, A. Zettl, and M. F. Crommie, Nat. Phys. **5**, 722 (2009).

[45] F. P. Toldin, J. Figgins, S. Kirchner, and D. K. Morr, Phys. Rev. B **88**, 081101 (2013); T. Yuan, J. Figgins, and D. K. Morr, *ibid.* **86**, 035129 (2012).

[46] J. Velev and W. Butler, J. Phys.: Condens. Matter **16**, R637 (2004).

[47] S. Katsura, T. Morita, S. Inawashiro, T. Horiguchi, and Y. Abe, J. Math. Phys. **12**, 892 (1971).

[48] A. J. Guttmann, J. Phys. A: Math. Theor. **43**, 305205 (2010).

[49] J. Cserti, Am. J. Phys. **68**, 896 (2000).

[50] T. Morita and T. Horiguchi, J. Math. Phys. **12**, 981 (1971); T. Morita, *ibid.* **12**, 1744 (1971).

[51] M. Berciu, J. Phys. A: Math. Theor. **42**, 395207 (2009); M. Berciu and A. M. Cook, Europhys. Lett. **92**, 40003 (2010).

[52] J. Pollmann and S. T. Pantelides, Phys. Rev. B **18**, 5524 (1978).

[53] R. Haydock, V. Heine, and M. Kelly, J. Phys. C: Solid State Phys. **5**, 2845 (1972); **8**, 2591 (1975); L. Falicov and F. Yndurain, *ibid.* **8**, 147 (1975); D. G. Pettifor and D. L. Weaire, *The Recursion Method and its Applications*, Springer Series in Solid State Sciences (Springer, Berlin, 1985), Vol. 58.

[54] For numerical calculations, poles are broadened using complex frequency $\omega + i\eta$ with finite η. For the QPI color plots, we use $\eta/t = 10^{-3}$, while for Brillouin zone cuts we use $\eta/t = 10^{-4}$ for improved accuracy near divergences.

[55] Specifically, T_K is defined in this work via $\mathrm{Im}\Gamma(\omega = T_K)\mathrm{Im}G_{\mathrm{d}}(\omega = T_K) = \frac{1}{2}\mathrm{Im}\Gamma(0)\mathrm{Im}G_{\mathrm{d}}(0)$.

[56] C.-Y. Lin, A. H. Castro Neto, and B. A. Jones, Phys. Rev. Lett. **97**, 156102 (2006).

[57] M. T. Glossop and D. E. Logan, Eur. Phys. J. B **13**, 513 (2000).

[58] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón, Phys. Rev. B **66**, 035412 (2002).

[59] T. O. Wehling, A. V. Balatsky, M. I. Katsnelson, A. I. Lichtenstein, and A. Rosch, Phys. Rev. B **81**, 115427 (2010).

[60] T. O. Wehling, A. I. Lichtenstein, and M. I. Katsnelson, Phys. Rev. B **84**, 235110 (2011).

[61] This property follows from Eqs. (26) and (27b), noting that $\boldsymbol{\tau} = \frac{1}{3}[\mathbf{a}_1 + \mathbf{a}_2]$ in terms of the real-space lattice vectors, and $\mathbf{a}_i \cdot \mathbf{A}_j = 2\pi\delta_{ij}$ as usual.

[62] M. T. Glossop and D. E. Logan, J. Phys.: Condens. Matter **14**, 6737 (2002).

[63] J. Mysliveček, A. Stróżecka, J. Steffl, P. Sobotík, I. Ošt'ádal, and B. Voigtländer, Phys. Rev. B **73**, 161302 (2006).

[64] L. Simon, F. Vonau, and D. Aubel, J. Phys.: Condens. Matter **19**, 355009 (2007).

[65] A. K. Mitchell, M. Becker, and R. Bulla, Phys. Rev. B **84**, 115120 (2011).