Probability of quasiparticle self-trapping due to localized energy deposition in nonequilbrium tunnel-junction detectors

Deborah Van Vechten and Kent S. Wood
Code 4120, U.S. Naval Research Laboratory, Washington, D.C. 20375-5000
(Received 22 August 1990; revised manuscript received 13 December 1990)

Voltage-biased, superconducting tunnel junctions are investigated as x-ray detectors for applications requiring both high quantum efficiency and better than 1% energy resolution. The nonequilibrium quasiparticles, produced as the energy deposited degrades to the few-meV-per-excitation level, tunnel and are detected before they are lost to recombination. Previous event modeling ignored the energy cascade under the assumption that the equilibrium of the electrodes is minimally perturbed by the deposited energy. In this paper we demonstrate that that assumption is invalid. We calculate the local energy density as the average quasiparticle energy becomes of the proper magnitude to suppress the gap. The fraction of the nonequilibrium quasiparticles that become spatially trapped (never to tunnel) in the order-parameter well that their existence creates may vary between events. If so, the source of the observed non-Poisson-limited energy resolution of this class of detectors would be identified. The input parameters used were evaluated in equilibrium. Thus our conclusions need to be confirmed via a fully nonequilibrium calculation of the cascade.

I. INTRODUCTION

Superconducting devices are currently under development as detectors for x- and γ-ray photons and energetic particles.[1] For this application, desirable performance characteristics include energy resolution of the order of 1–10 eV for 6 keV photons incident and at least a 10% quantum efficiency. This paper focuses on Josephson (or Giaever) tunnel junctions (TJDs) that directly sense perturbations in the single-particle tunneling current at a bias voltage near $V=\Delta_{=}/e$. Unlike pulses seen in junctions biased at the gap ($V=2\Delta_{=}/e$), pulses in these junctions are normally asserted to result solely from the short-lived, dramatic increase in the number of quasiparticles (qp's) present in the device. (The volume within which equilibrium has been disturbed by the x-ray energy will sometimes be simply referred to as the "disturbed volume".)

Various groups have developed TJDs using granular tin tunnel junctions[2] and have achieved energy resolutions near 0.7%. Conceptually identical work using niobium junctions, which thermally cycle much more reliably, is in progress.[3] To date, 84 eV is the best reported resolution in niobium,[3(a)] which, while less good than in tin, is already better than in cooled semiconductor detectors which sense electron-hole production. The Poisson limit on the resolution of the superconducting detectors is less than 10 eV out of 6 keV incident.

Understanding the process by which the photon's energy is read out by the junction detector should contribute to device optimization. To model the process comprehensively, division of the event into three distinct stages is desirable. Stage 1 consists of the initial photoelectron collision cascade, wherein the most likely energy per excitation rapidly degrades to the eV level. Stage 2 encompasses the subsequent expansion of the disturbed volume and continued degradation of the energy. By definition, it ends when a typical excitation energy is of the order of the equilibrium superconducting pairing energy $2\Delta_{=}$. Stage 3 includes the remainder of the expansion and energy degradation during which thermal equilibrium is established first locally within the sample and then with the bath.

Published papers on the tin junction experiments[4] generally include results from Rothwarf-Taylor modeling of the pulse shape. These calculations correspond essentially to stage 3 modeling under the assumption that stage 2 concludes with the local gap undisturbed. This modeling is fully successful at matching the pulse shape. Twelve plausibly valued materials parameters (e.g., phonon lifetime against pair breaking and anharmonic decay, qp recombination probability, and propagation velocity) and fitting constants (e.g., qp and phonon transparency of each layer interface) are used. However, success in reproducing the pulse shape does not guarantee that the initial assumptions and the underlying model are correct. Indeed, probing the details of the energy degradation cascade experimentally is very difficult. For integral detectors, the output electronics smear the temporal details of the pulse and mainly leave information on the excess number of carriers that tunneled, on the variance in that number for monoenergetic input, and on the pulse rise time. Distributed detectors[5] have only one additional output parameter, namely the variation in signal strength with distance between the event and readout junction, that distance being itself inferred from the data modeling.

Other theoretical models of the event sequence initiated by photon capture may equally well explain the experimental pulse shapes. The present paper begins to explore this possibility by analyzing aspects of the energy deposition and re-equilibration processes that could limit the energy resolutions. Both processes are currently without

a completely satisfactory description, in large part because of involvement of unresolved problems from other areas of physics and materials science research. Thus the major goals of this paper are (i) to call attention to these adverse factors, (ii) to constrain these resolution degrading factors quantitatively, and (iii) to identify the main obstacles that must be overcome in order to arrive at better models.

Section II will consider stage 1 and the "initial" energy density produced by photon deposition. This exceeds by many orders of magnitude the energy density needed later in the process to cause the disturbed region to transition to the normal state. Section III treats the evolution of the disturbed volume through stage 2. Estimates are given for its ending volume, the stage's duration, the diffusive expansion factor, and the rates at which the energy distribution evolves. From these estimates we infer that two kinds of loss enhancing processes—quasiparticle self-trapping and phonon bottlenecking—are likely. The fraction typically lost will depend on the end of stage 1 energy density and thus on the dE/dx of the incident particle. Thus the linearity of the device response with incident energy and comparison of the response to pulses of the same total energy but differing numbers of photons should probe the inherently nonequilibrium loss mechanisms. They may also limit the energy resolution of monochromatic x rays. Variation in the detailed evolving energy distributions due to variations in the energy transferred in each collision can produce variation δN_i from event to event in the number N_i of quasiparticles available to diffuse and tunnel, and so to be detected. The energy resolution will never improve beyond $\delta N_i/N_i$. This point and related topics concerning laboratory work now in progress are discussed in Sec. IV.

II. EVOLUTION OF THE STAGE 1 CASCADE: DISTURBED VOLUMES AND ENERGY DENSITIES

When energy deposition occurs by x-ray photon absorption, essentially all the incident energy eventually goes into either the electronic or phonon system. Photons produce few nuclear displacements until the incident energy exceeds 100 keV because all the energetic particles moving through the lattice are electrons. Heavier, more highly ionizing incident particles such as α particles have an order of magnitude shorter range and may lose significant energy to nuclear damage. Other channels for energy storage and degradation, for example plasmons or perturbations of the local nuclear magnetic moment configuration, undoubtedly exist. However, nonexistent calculations of, for example, plasmon lifetimes in metals near $T=0$ are needed to constrain such possibilities. Thus we defer them to a later paper. The evolution of the energy initially imparted to the photoelectron is our focus.

It is important to estimate the spatial distribution of the energy at each stage of the energy degradation process. The time delay required for the energy to diffuse from the event site to the tunneling barrier in stage 3 is not the only consequence of this distribution. The spatial distribution at the end of stage 1 is quite well studied because of its relationship to doping profiles after ion implantation. When the photon's energy is transferred to the initial photoelectron, the interaction takes place in an atomic volume. As the resultant primary photoelectron exits, it immediately begins to scatter inelastically. These scatterings produce secondary electrons and cause the primary's path to become irregularly kinked. The paths of the secondaries branching off from that of the primary are also highly irregular, since they also reflect a sequence of random scattering events.

Stage 1 is *defined* to be over when the primary has lost enough energy to become indistinguishable from the other electrons in the system, i.e., when it possesses of the order of a few eV of energy. The distance the electron will have traveled in losing the rest of its initial energy is called its "range." While the ranges associated with individual events may vary by factors of 2, averaged over many events a statistically significant value R_e is determined. For photoelectrons with energies above 10 keV, the range R_e in microns is experimentally found to be[6]

$$R_e(E)=\frac{3.52}{\rho}\left[\frac{E}{10\ \text{keV}}\right]^{1.754}, \tag{1}$$

where ρ is the density of the material in g/cm^3. Equation (1) can also be used to estimate the distance a secondary of energy E travels before its kinetic energy drops to a few eV. At the end of stage one, there will be substantial numbers of excitations with energies less than an eV corresponding to the decay products of lower energy excitations produced earlier in the cascade. Ion implantation modeling indicates that stage 1 takes of the order of a 1 ps or less for massive incident particles of hundred keV or more energy. A bound of 10^{-14} s can also be derived by assuming a uniform deceleration to zero velocity of an electron with an initial kinetic energy of 10 keV.

To proceed with a quantitative calculation, we need an estimate of the volume in which the energy resides at the end of stage 1. The heavily branched, irregularity bending "track" of scattering events has a volume that is essentially the product of the primary electron range with an effective cross-sectional area. We thus treat the track as a cylinder of length equal to R_e. The range of a secondary of "typical" energy is used as the cylinder radius, even though the actual energies of secondaries are expected to vary by at least a factor of 10 in a real event. We arbitrarily define 20 secondaries as "typical" (more are likely) and equally partition the energy. This procedure *overestimates* the affected volume because 20 secondary arms arrayed perpendicular to the axis cannot possibly full the cylindrical volume. Substitution of a 6 keV photoelectron energy into Eq. (1) predicts the primary electron range to be 0.17 μm for Nb and 0.25 μm for Sn, while Eq. (1) yields 0.89 nm for Nb and 1.3 nm for Sn for the secondaries' range.

Thus at the end of stage 1, the energy will have degraded to the $\sim$ eV/quantum level within a volume of 4.2×10^{-7} μm^3 for Nb and 1.4×10^{-6} μm^3 for Sn. If the entire incident photon energy is to be correctly measured by the device, the junction must be thick enough to contain the track.[7] This establishes a minimum junction

thickness of R_e. A quantum efficiency of 25% for absorption of 6 keV x rays requires the greater thicknesses of 1.05 and 0.75 μm for Nb and Sn, respectively. Achieving high quantum efficiencies generally guarantees that few photoelectrons will escape.

In time, the boundary of the disturbed volume expands and the local energy density within it drops. Simultaneously, both the maximum and the average energy per excitation decreases. The evolution until the mean excitation energy degrades to a few meV is designated as stage 2. The published models based on the Rothwarf-Taylor equations[4] begin with stage 3 and make several assumptions in establishing the initial conditions. The most important assumption is that the modeling can begin with all the electronic energy residing in gap edge qp. As we shall demonstrate, the validity of this assumption requires that thermal equilibrium be only weakly disturbed at the end of stage 2. However, it is far from obvious that there is sufficient time during stage 2 for these starting assumptions to be valid. It is quite possible the details of stage 2 remain evident when the pulse of quasiparticles is collected during stage 3.

We first consider the energy density in the disturbed volume at the end of stage 1. Using the volumes estimated above, this density is found to be at least 1.4×10^{10} eV/μm^3 for Nb and 4.4×10^{9} eV/μm^3 for Sn. On the energy scale of the superconductivity, these energy densities are enormously high. By using conventional estimates of the Cooper pair density,[7] the condensation energy associated with breaking all the pairs that have *any* part of their wave functions within the end of stage 1 volume is found to be only 40 meV for Nb and 3 meV for Sn. That is, the energy density at the beginning of stage 2 exceeds the condensation energy of the pairs "in" the same volume by a factor of 10^5 for Nb and of 10^6 for Sn. Table I lists values of the expansion factors needed during stage 2 for all the elemental superconductors. Substantial expansion must occur during stage 2 if the disturbed volume is not to be driven normal.

III. STAGE 2: EVOLUTION OF THE ENERGY DISTRIBUTION

The quasiparticles present at the end of stage 2 tunnel to create the pulse detected during stage 3. Hence the production and evolution of the qp population are at the heart of detector performance. The number and energy distribution of the qp evolves markedly during stage 2. Thus the effect of that distribution on the superconducting order parameter, Δ, is time dependent. At the beginning of stage 2, the most energetic electronic excitations[9]

TABLE I. Material dependent parameters (Ref. 8) of the elemental superconductors. Given are the bulk zero field transition temperature, T_{c0}; zero temperature thermodynamic critical field, H_0: Debye energy, Θ_D; mass density ρ; and atomic number Z. R_e and R_s are the ranges of the primary and "typical" secondary photoelectrons, calculated from Eq. (1). D_1 is the estimated minimum energy density at the end of stage 1 of the cascade. D_H is the condensation energy density of the pairs. The parameter f_D, written in the form prefactor (power of 10), is the minimum dilution factor needed during stage 2 to prevent the disturbed volume from going normal. Lastly, r_2 is the radius of the smallest sphere which could contain 6 keV and not go normal. In a spherical ($R_e=0$) approximation, it is the expansion needed during stage 2.

	T_{c0} (K)	H_0 (G)	Θ_D (meV)	ρ (g/cm^3)	Z	R_e (6 keV) (μm)	R_s (nm)	D_1 (10^9 eV/μm^3)	D_H (eV/μm^3)	f_D	r_2 (μm)
Al	1.175	104	36.2	2.69	13	0.534	2.80	0.46	270.000	6[−7]	1.70
Be	0.026	1	120.0	1.84	4	0.781	4.10	0.15	0.025	2[−10]	37.00
Cd	0.517	28	18.0	8.65	48	0.166	0.87	15.00	20.000	1[−9]	4.10
Hf	0.128	13	21.7	13.31	72	0.108	0.56	55.00	4.200	8[−11]	6.90
Hg	4.150	411	6.9	13.54	80	0.106	0.55	60.00	4200.000	7[−8]	0.70
In	3.408	281	9.4	7.31	49	0.197	1.00	9.10	2000.000	2[−7]	0.90
Ir	0.113	16	36.6	22.42	77	0.064	0.34	260.00	6.400	3[−11]	6.00
Mo	0.915	96	39.7	10.22	42	0.141	0.73	25.00	230.000	9[−9]	1.80
Nb	9.250	2060	23.8	8.57	41	0.168	0.88	15.00	110000.000	7[−6]	0.24
Os	0.660	70	43.1	22.57	76	0.064	0.33	270.00	120.000	4[−10]	2.30
Pb	7.196	803	8.3	11.35	82	0.127	0.66	35.00	16000.000	5[−7]	0.45
Re	1.697	200	35.8	21.02	75	0.068	0.36	220.00	990.000	5[−9]	1.10
Ru	0.490	69	50.0	12.41	44	0.116	0.61	44.00	120.000	3[−9]	2.30
Sn	3.722	305	16.8	5.75	50	0.250	1.30	4.50	2300.000	5[−7]	0.86
Ta	4.470	829	22.2	16.65	73	0.086	0.45	110.00	17000.000	2[−6]	0.44
Tc	7.800	1410	35.4	11.50	43	0.125	0.65	36.00	49000.000	1[−6]	0.31
Th	1.380	160	14.2	11.72	90	0.123	0.64	38.00	640.000	2[−8]	1.30
Ti	0.400	56	35.8	4.54	22	0.316	1.70	2.20	78.000	4[−8]	2.60
Tl	2.380	178	6.8	11.85	81	0.121	0.63	40.00	790.000	2[−8]	1.20
V	5.400	1408	33.0	6.11	23	0.235	1.20	5.40	49000.000	9[−6]	0.31
W	0.015	1	33.0	19.30	74	0.074	0.39	170.00	0.025	1[−13]	37.00
Zn	0.850	54	26.7	7.13	30	0.202	1.10	8.60	72.000	8[−9]	2.70
Zr	0.610	47	25.0	6.50	40	0.221	1.20	6.40	55.000	9[−9]	2.90

have energies of a few eV. The $\Delta(t)$ must at all times be calculated self-consistently from

$$\ln[\Delta(0)/\Delta(t)] = \int dE[\rho(E)f(E,t)/E] , \quad (2)$$

where $\rho(E)$ is the density of states of the quasiparticles and $f(E,t)$ is the time-dependent probability of their occupation. Since $\Delta_=$ is of order 1 meV, the excitations characteristic of the beginning of stage 2 are of too high an energy to have any effect on Δ. However, by definition of the end of stage 2, the average excitation energy of the excess (nonequilibrium) qp will then be of order $\Delta_=$. Such excitations can have a substantial effect on Δ. Moreover, their absolute number will have become large ($\geq 10^6$), since energy is conserved. *Unless* the volume of the disturbed region has greatly enlarged during stage 2, the local $f(E,t)$ function will also have become large near $\Delta_=$ and severe suppression of the order parameter, possibly to zero, will result.

Figure 1 portrays the expansion during stage 2. Surface "*a*" describes the basic cylinder approximation to the illustrated photoelectron track volume and so represents the volume at the beginning of stage 2 (4×10^{-7} μm^3 for Nb). Surface "*b*" represents the maximum volume (5.7×10^{-2} μm^3 for Nb) guaranteed by thermodynamics to have been driven normal at the end of stage 2. Surface "*c*" corresponds to the minimum volume (18 μm^3 for Nb) the energy may occupy at the end of stage 2 for the order parameter to be suppressed by less than 1%. As we shall demonstrate, any suppression of the order parameter creates the possibility of locally trapping the nonequilibrium quasiparticles and violates the assumptions of Rothwarf-Taylor modeling. Thus, surface *c* represents the minimum volume the photon energy may occupy for Rothwarf-Taylor modeling to be strictly valid.

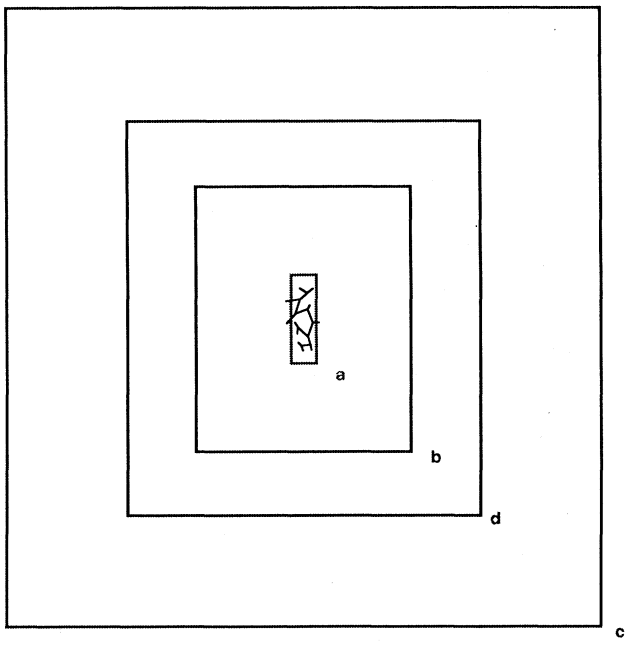

FIG. 1. Schematic description of the stage 2 expansion. The innermost surface "*a*" represents the cylinder approximation to the photoelectron track, the end of stage 1 volume equalling 4×10^{-7} μm^3 in Nb. Surface "*b*" represents the minimum end of stage 2 volume to not go fully normal, 5.7×10^{-2} μm^3 in Nb. Surface "*c*" is the minimum volume for less than 1% gap suppression, 18.0 μm^3 in Nb if all the excess energy goes into quasiparticles. Surface "*d*" represents the result of the equilibrium system time estimate of the stage 2 expansion, 0.2 μm^3 in Nb.

Accurate event modeling is thus seen to require accurate estimates of the expansion time and rates during stage 2. While some details of the stage 2 energy degradation processes which influence the duration are difficult to determine, the broad outline is clear. Figure 2 displays the dispersion curve of a typical superconductor, where the energy gap is significantly exaggerated in order that it be visible. Note that there is an absolute minimum at the local value of Δ at the Fermi momentum. Further, recall that the density of qp states has an intense spike just before it drops to zero at Δ, so that the Pauli principle does not prevent relaxation of the excess quasiparticles into minimum energy states. Once the quasiparticles relax to that energy minimum, only two loss mechanisms are available. The quasiparticles can physically migrate out of the device volume. Alternately, they can effectively migrate out by recombining into pairs by emitting phonons and have the phonons either escape into the leads or substrate or decay into lower energy phonons without breaking additional pairs. (If the recombination phonons do neither of these, they will break some pair, thereby restoring the initial number of quasiparticles.) Higher energy quasiparticles have a shorter effective lifetime. They can also disappear from their initial energy bins by relaxing "down the dispersion curve" through phonon emission.

Quasiparticle lifetimes due to energy decay are quantified in Fig. 3. It plots the $T=0$ limit of the energy

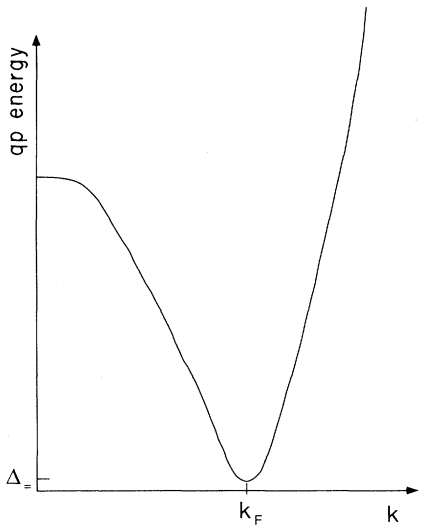

FIG. 2. Schematic representation of the quasiparticle dispersion curve with exaggerated gap.

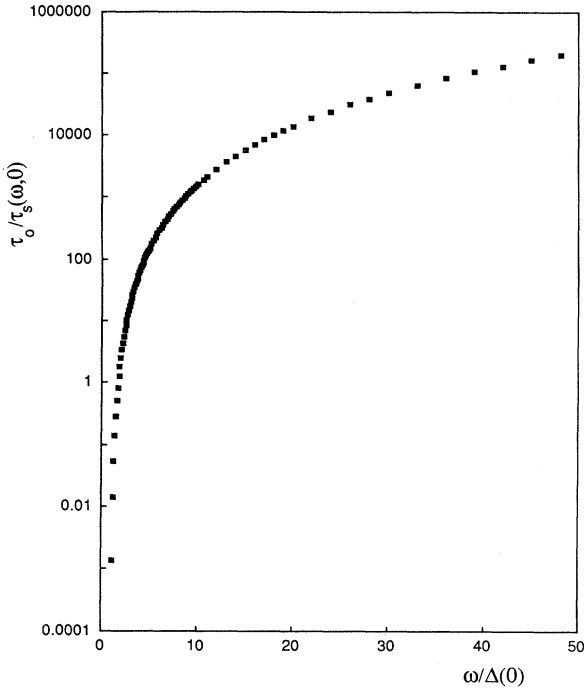

FIG. 3. The lifetime against decay of qp energy by phonon emission as a function of quasiparticle energy under the Debye approximation at $T=0$. Graph is a plot of Eq. (17) of Ref. 10. For Nb, the τ_0 parameter is 0.149 ns.

dependence[10] of this decay (relaxation) lifetime time τ_s under the approximation that the Debye phonon energy relation $E \propto b\omega^2$ applies to all energies of interest. The plot shows that the lifetime against phonon emission of an excitation is five orders of magnitude longer at 1.2 than at 9.6 times the local gap. Indeed, at 4Δ in Nb, τ_s is 2.2 ps, while it lengthens to 59 ps at 2Δ. In contrast, for energies above 33.5 times the gap, the lifetime against phonon emission is less than 10 femtoseconds in Nb. We conclude that it is the duration of the last few decay generations that effectively determines the time scale of stage 2. Moreover, the width of the qp energy distribution, (if it is more than $\sim 0.3\Delta$ wide when the peak is less than $\sim 4\Delta$) may well impact the average timescale.

The quasiparticles produced by decay of higher energy electronic excitations, such as core-level holes and electrons near the top of the conduction band, eventually accumulate at the local Δ. If at that time the total energy density in the local region is consistent with thermal equilibrium at the bath temperature, then the absolute minimum possible excitation energy will be $\Delta_{=}$. Quasiparticles decaying to this energy would still be free to propagate throughout the electrode. However, if the local gap is suppressed to Δ by the presence of the nonequilibrium quasiparticles, this dispersion curve minimum will occur at Δ, not $\Delta_{=}$. No physical process would then exist to differentiate $\Delta_{=}$ from any other excited state quasiparticle energy or cause it to be preferentially populated. Indeed, consider the possibility that a significant number of quasiparticles, resulting from a sequence strictly of inelastic scattering events (no recombination involved) arrive at the dispersion-curve minimum noticeably earlier than the other qp. The early arrivals will act to suppress the local gap. This would increase the decay rates for the qp still at higher energies. We speculate that an avalanche effect can occur which sweeps the bulk of the qp through $\Delta_{=}$ to lower energies.

Not all the qp in the electrodes at the end of stage 2 originate from the decay of high-energy electronic excitations. Additional minimum energy quasiparticles are formed via direct pair breaking by phonons with energy above twice the local gap. (According to Ref. 9, for fixed phonon energy, the phonon pair breaking rate is roughly constant for temperatures T below $T/T_c < 0.4$. The rate increases with phonon energy slightly less rapidly than linearly.) Phonons with energies in the $2-4\Delta$ energy range will be especially efficient at producing qp with energies within Δ of the local gap edge. These quasiparticles will have the largest possibility of affecting the local gap and are also relatively slow to decay to the gap energy. The production rate of such qp will be highest during the time interval in the energy decay when this phonon energy component is the most densely populated.

Thus the time evolution of the phonon energy distribution must also be considered. Those phonons of energy $<2\Delta$ emitted as the qp relax toward Δ are ineffective as pair breakers unless they enter a region of yet lower gap where they are no longer "subgap." Thus they are normally ignored as irrelevant in Rothwarf-Taylor modeling. However, when these phonons have a high local number density, the rate of their reabsorption by quasiparticles is enhanced, thereby lengthening τ_s. Phonons of energy $>4\Delta$ can effectively relax by first pair breaking, then having the resultant qp emit lower energy phonons and recombine. Thus there are a wide variety of ways to transfer energy into the electronic energy system from the phonon system. Indeed, the point in time when the partition of energy between the electronic and phonon systems is finalized does not occur before the end of stage 2. We suggest that the partition is finalized at the point in stage 3 when the mean phonon energy becomes subgap and energy transfer from the phonon to the quasiparticles becomes improbable. Only then does qp recombination represent a true loss term.

When $\Delta < \Delta_{=}$, all the nonequilibrium qp in the disturbed volume whose energies degrade to the bottom of the local well[11] are effectively spatially trapped there. Their energies correspond to virtual states in the remainder of the electrode. Thus propagation outside the well is prohibited by energy conservation and the states are spatially localized.[12] (Put another way, Andreev reflection traps qp having energies less than $\Delta_{=}$, just as it ensures an *N-S* interface has a high thermal boundary resistance.) Those quasiparticles in the ensemble which manage to diffuse into the region outside surface *c* in Fig. 1 before their energies degrade to $\Delta_{=}$ are unlikely to be trapped because their local density will be too low to cause a well to form. However, the time scale estimates indicate that it is much more likely the primary question is whether the qp's get further out than surface *b*, so that the disturbed volume does not go normal, before this

time.

We have identified only three mechanisms whereby the depth of the order-parameter well will ever decay: (1) recombination of the qp and the escape of the resultant $<2\Delta_{=}$ phonon into the region outside the well where it is unable to break pairs; (2) excitation of trapped quasiparticles by phonon absorption to energies $>\Delta_{=}$ so that escape is possible; and (3) entry of a net pair current into the region, driven by the gradient in Δ. Mechanism 1 transfers energy from the qp to phonons. Its impact on energy resolution will be discussed shortly. Mechanism 3 is suggested by the observation of induced gaps in normal metals in the time-independent (equilibrium) proximity effect. However, for such pair currents to have an effect, occupied allowed qp states localized within the well must be depopulated because their energies are forbidden once the well heals. No mechanism is obvious to us for this to occur except some sort of dilution of the qp by excess pairs, and that would violate charge neutrality. The time scales and significance of these possibilities will be evaluated in a later paper.

A. Estimates of the duration of and expansion during stage 2

The question of how much expansion the disturbed volume experiences during stage 2 is obviously tied to the question of how long the expansion lasts. Gray has estimated[13] the time for a representative excitation to decay in energy from 100Δ (roughly the Debye energy, i.e., the assumed maximum phonon energy) to Δ. That the energy per excitation of the qp and phonons decays at the same rate, essentially because the two populations are in mutual equilibrium, is assumed. Thus the component with the more slowly decaying scale defines the time scale of the entire decay. Five decay generations are, on average, required to drop the average energy from the Debye energy to Δ. Applied to Nb, phonon relaxation determines the time scale and stage 2 is asserted to require 10 ps. (Since the Debye energy of Nb is only 17Δ, this ignores 1.5 generations. However, all the missing generations are of such a high energy that ignoring their durations introduces little error.)

As Gray points out, a diffusion constant characterizing the entire 1 eV to 1 meV cascade of 25 cm^2/s would combine with this 10 ps time estimate to predict an outward qp diffusion of 0.2 μm in Nb. (We remind the reader that many interactions, each one of which causes memory of the previous direction of propagation to be completely lost, occur during this time. One cannot estimate the expansion occurring during stage 2 by multiplying its 10 ps duration by the qp velocity.) A recently measured[14] value of the quasiparticle diffusion constant D during stage 3 in bulk Nb is 60 cm^2/s. Using this value, instead of Gray's, would increase the predicted diffusion to 0.3 μm.

Use of the stage 3 value for D guarantees that 0.3 μm *overestimates* the stage 2 expansion for Nb. To see this, note that the diffusion constant D is linear in the mean-free path between collisions λ. λ depends on, for example, the phonon emission probability, as well as the probability of elastic scattering. Since τ_s is less than 1 ps at energies above 5Δ, the λ which characterizes the early part of stage 2 will be substantially shorter than that characterizing the gap edge qp propagating in an undisturbed region of the electrode. (The latter is often approximated by the mean-free path of thermal quasiparticles propagating in the sample in the normal state. While this practice is usually the only easy alternative, the lower average excitation energy of the normal state quasiparticles ($<\Delta_{=}$) makes it suspect.)

Estimating the stage 2 expansion is, unfortunately, not actually this cut and dried. The assumption in Gray's time scale estimate that the phonons and qp are everywhere in mutual equilibrium is questionable. Reference 15 reports that stage 3 qp and phonons decouple below 3 K in single crystal lead. (The qp pulse arrives first.) By 2 K, decoupling has also occurred in single crystal Nb, illuminated on the back by visible light. Here the phonon pulse arrives before the quasiparticle pulse.[14] The requirement of a low thermal quasiparticle density in the tunnel junction detectors causes a typical operating temperature to be below 0.1 of T_c or 0.9 K for Nb. Thus if stage 2 ends with the disturbed volume essentially in thermal equilibrium with the bath, the phonons and quasiparticles will have decoupled before that end is reached.

We choose to ignore these questions of when the qp and phonons are in mutual equilibrium in order to pursue questions concerning the final outcome. The cylindrical approximation for the disturbed volume and energy density at the beginning of stage 2 can be used to estimate the stage 2 final volume. Enlarging the cylinder in all directions by 0.3 μm, our estimate of maximum expansion during stage 2 in Nb, gives an upper bound on the disturbed volume of 0.2 μm^3 and a minimum energy density of 3×10^4 $eV/\mu m^3$. The latter is a third of the Nb pair condensation energy. Surface "*d*" in Fig. 1 schematically represents this 0.2 μm^3 volume. Applying the same estimation procedure to the much weaker electron-phonon coupling case of aluminum, Gray[13] predicted a 2 ns duration for stage 2 and two microns of diffusive enlargement of the affected region. Thus the Al 5×10^8 $eV/\mu m^3$ energy density at the end of stage 1 is predicted by our cylindrical geometry argument to decrease to 106 $eV/\mu m^3$ by the end of stage 2. This is approximately 40% of the Al condensation energy. While not high enough to guarantee that the disturbed volume is driven fully normal, these estimates strongly suggest the order parameter is suppressed as the excitations degrade down to the equilibrium gap edge.

A second calculation exists in the literature for the duration of stage 2, this time in lead.[16] It is somewhat more exact in that it uses the measured phonon spectra, rather than the highly simplified Debye spectrum. This calculation assumes the qp and phonon energy distributions are independent and deals only with the quasiparticles. Thus it follows the mean excitation energy down from an eV by first considering screened electron-electron scattering and then phonon emission by the highly excited qp. A total time of 1 ps is estimated for stage 2 during which 20 secondary quasiparticles and 200–400 phonons are generated for each initial 1 eV qp. Individual scattering/decay events are less than 25 femtoseconds

apart. Even assuming the qp propagate at 10^8 cm/s, this corresponds to a maximum mean-free path of 2.5×10^{-2} μm. Since $D=60$ cm^2/s corresponds to a mean-free path of 0.1 μm, this time estimate corresponds to a prediction of a smaller amount of outward diffusion during stage 2 than does Gray's. Since that 10 ps, 0.3 μm estimate was too short for the gap to avoid being locally suppressed, this shorter estimate corresponds to more severe suppression, possibly to the point of phase transition.

There is one experimental measurement that relates directly to the question of the duration of stage 2, namely a pump/probe measurement[17] of the changes in the reflectivity of large grained $YBa_2Cu_3O_{7-\delta}$ films. With a time resolution of 60 femtoseconds, the reflectivity was observed to drop for 300 femtoseconds followed by a 3 ps relaxation to an enhanced value characteristic of the bolometric response. The first phase is interpreted[17] as corresponding to the decay of the energy of highly excited electrons into optical phonons and subsequent Cooper pair breaking. The later phase is interpreted[17] as indicating the predominance of quasiparticle recombination.

B. Significance of nonequilibrium nature of the disturbance

The above arguments clearly indicate that one should expect the gap to be significantly suppressed at the beginning of stage 3. The only reason to question this conclusion is that both stage 2 duration estimates use as input the energy decay rates calculated by assuming only one excitation is present at a time. This corresponds to the response of the equilibrium system. However, at the end of stage 1, the system is locally extremely out of equilibrium. Thus the assumptions of the estimates are inconsistent with the situation to which they are applied. In order to determine the extent of gap suppression, consideration of the disturbed nature of the system is essential.

To see the validity of this assertion, consider the phonons at the end of state 1. The local energy density (10^8 to 10^{11} eV/μm^3) is so high that phonon bottle necking is likely to occur. When each incident photon carries 1 eV, a typical excitation energy at the end of stage 1, phonon bottle necking occurs[18] in hydrated amorphous silicon if an IR pulse deposits $\sim10^6$ eV/μm^3. (Evidence of local melting has been observed in crystalline Ge at lower energy densities than the superconductors experience at the beginning of stage 2.) In the x-ray case, branching ratio calculations starting at 100Δ indicate the quasiparticles carry less than 40% of the total energy.[19] This still leaves the local phonon energy density sufficient to induce bottle necking during a portion of stage 2, especially when more heavily ionizing (larger end of stage 1 energy densities) particles are incident.

In phonon bottle necking, the phonons are produced with a wide range of energies and therefore group velocities. (v_g is proportional to the slope of the dispersion curve, hence v_g tends toward zero at the Brillouin zone edge.) Enough "high" energy phonons with small group velocities are locally produced that even the higher group velocity, low-energy phonons are heavily scattered. This slows their escape, effectively impeding the outward diffusion of the energy. The presence of a large density of excitations causes the behavior of each one to be modified, when compared to the behavior it would have had in the equilibrium system.

Exactly how significantly phonon hot-spot formation lengthens the duration of stage 2 is a question beyond the scope of this discussion. However, it is clear that hot-spot formation will have consequences. Even in a mild hot spot, only the low-energy phonons are able to diffuse away. Radial spatial variations in the phonon energy distribution are thus likely to develop. Pair breaking by the escaping phonons will then transfer this radial variation to the quasiparticle distribution, since it is an important qp production mechanism. Phonon focusing along crystallographically preferred directions in large grained (or textured) deposits could similarly produce angular variations. (The random crystallographic orientation of adjacent grains in fine-grained, polycrystalline material tend to average out this effect.)

IV. COMMENTS AND CONCLUSIONS

The energy density at the end of stage 1 far exceeds the equilibrium superconducting pair condensation energy density. Equilibrium-based estimates of stage 2 duration and the extent of stage 2 expansion indicate that suppression of the local order parameter will occur in the disturbed region. Moreover, since the time scale of the decay of the qp energy $E/(dE/dt)$ is of the same order of magnitude as the time required to diffuse only to a large enough radius that the local gap has a chance not to vanish, some qp trapping in the well is inevitable.

Our conclusion that localized gap suppression is likely is in fact supported by the Rothwarf-Taylor modeling of the expansion of the disturbed volume in Ref. 1(b). The paper by W. Rothmund and A. Zehnder contains a plot of the spatial dependence of the suppression of the local value of the order parameter of tin for 0.11 and 0.3 ns elapsed after the x-ray event. These calculations assume that all the electronic energy is in gap edge qp and set the qp diffusion constant equal to the normal state value. These choices tend to minimize the effect of the locally elevated energy density and most closely approximates the experimental situation during stage 3. Nevertheless, the plot shows an 18% suppression of the local order parameter after 0.3 ns. [For bulk materials $\Delta(t)\geq0.95\Delta(0)$ for all reduced temperatures t below 0.51.] Thus this plot indicates that gap suppression occurs and lasts longer than 0.3 ns even if all the details of the preceding energy cascade are ignored and no gap suppression is assumed to exist when the qp are formed.

We noted earlier that it is customary in Rothwarf-Taylor modeling to use a single time and energy independent qp diffusion constant D. Once the expansion of the disturbed volume has proceeded far enough that the qp energies are not changing on the timescale of the propagation and all regions of the sample are equivalent, this practice is fully appropriate. However, it is tempting to try to use the Rothwarf-Taylor approach to describe the entire x-ray event.[20] Given the rapid changes in energies and densities of qp during stage 2, such modeling can be

expected to work well only if the quasiparticle mean-free paths $\lambda(t)$, pair breaking lifetimes $\tau_B(t)$, and qp recombination rates used are appropriately averaged over the space and time dependent energy distributions. In addition, during the time interval when there is an order-parameter well, spatial dependence is expected in the effective $\lambda(t)$. qp trapping can occur and the relatively high local qp density within the well enhances the qp-qp scattering rate. Both effects shrink the effective mean-free path in the vicinity of the well, the former in an energy-dependent way. In the presence of an order-parameter well, the threshold for a given phonon to pair break becomes spatially dependent. It would seem more accurate to divide phase space into many cells, and allow each cell to evolve independently during each time step.

It is highly desirable that the response of a detector be linear in the incident energy. If stage 2 expansion is so extensive that the equilibrium is only weakly disturbed and the local gap remains $\Delta_=$, then both the density of fully relaxed nonequilibrium quasiparticles at the end of stage 2 and the resulting amplitude of the current pulse will be linear functions of the incident energy. However, if the qp's that tunnel are created in regions of differing Δ, the region of origin will influence the energy cost ϵ of each individual qp which tunnels and therefore the effective ϵ of the ensemble which tunnel. Even if this ϵ variation is somehow avoided, ϵ will depend on Δ and therefore the degree of order-parameter suppression. Comparisons of the degree of linearity of the pulse amplitude with total pulse energy as a function of dE/dx of the initial ionizing particle are thus expected to test unambiguously the degree of disturbance of Δ from $\Delta_=$ at the end of stage 2. Linearity of this sort has not yet been demonstrated experimentally. Indeed, the only information—an observed 21% discrepancy in the measured position of the $K\beta$ peak in the tin junctions (compared to its position as predicted from the $K\alpha$ peak and assumed linearity from $E=0$ to the $K\alpha$ energy, with zero offset at the origin)—suggests nonlinearity as a function of total energy. More complete tests of linearity in incident energy are underway[21] using visible light pulses of varying numbers of photons arriving over a uniform time interval. Comparison to the responses associated with ^{55}Fe x-ray photons and α particles of the same total energies as the visible pulses is also planned.

In addition to the question of linearity, gap suppression may directly impact the energy resolution. In any detector where energy is measured by counting quanta produced at energy cost ϵ per quantum, it is possible for the resolution to be dominated by Poisson statistics. In that case, the number of quanta is $N_i=E_\gamma/\epsilon$ and the resolution is $\delta E/E=N_i^{-1/2}$. This is the best possible resolution in such a detector. A larger value of $\delta E/E$ will be measured if there is some intrinsic scatter from event to event in the expectation value for N_i, e.g., if ϵ varies with position in the detector. Something like this is currently being observed in tin tunnel junction detectors. From calibrations of pulse heights it is known that roughly 1.2×10^6 qp are being collected each event, but the resolution is 41 eV for 6 keV incident photons. Poisson statistics on the observed qp count would permit much better resolution, roughly 6 eV. There is thus a practical question: why does observed resolution fail to reach the Poisson-limited value for the number of qp inferred from the pulse height?

Our paper has examined a mechanism—self-trapping—by which N_i could be smaller than predicted from the equilibrium gap energy and the equilibrium qp-phonon branching ratio and the incident energy. We have described the conditions under which it will occur in perfectly homogeneous material. Spatial inhomogeneities which act to shorten either the qp or phonon mean-free path will increase the likelihood of self-trapping. The mechanism has the capacity to reduce the pulse amplitude by factors of 50% or more. Additionally, degradation of the energy resolution (below the Poisson-limited estimate) would result from event-to-event variation in the number of quasiparticles self-trapped at the end of stage 2.

We are not attempting to assert that self-trapping is the only contribution to the observed non-Poisson limited resolution. The degradation in resolution from roughly 6 to 41 eV can also be the result of any sort of irregularity or nonuniformity that varies the production and collection efficiency of qp from event to event. For example, nonuniformity of tunneling barriers or localized trapping sites in the tunnel barrier or electrode could contribute to such variations. The idea that self-trapping considerations might dominate resolution needs further investigation, either by detailed calculation of the trapped fraction or by experiments that lead to inferences about how much trapping occurs in stage 2. The linearity experiments described above should shed some light on this question.

If the disturbed volume actually does go normal as the energy cascades through the meV-per-excitation level, the central image of how the nonequilibrium quasiparticle tunneling detectors operates must be modified. In particular, the excess qp that are observed to tunnel would arise as the track volume is being restored to thermal equilibrium at the bath temperature from the normal state. The number of qp that tunnel would reflect the rate of energy loss to the substrate and leads, and the exponential thermal qp density determined by the local temperature at the boundary of the disturbed volume when it expands enough to intersect the tunneling barrier. Linearity of the overall detector's response would depend on the material's heat capacity being linear in the maximum local temperature reached. The ultimate energy resolution would depend on effects that cause variation in the effective temperature at the onset of tunneling. These effects could include the distance from the event site to the barrier, the positions of internal grain boundaries, and the thickness of the electrodes.

V. SUMMARY

In the above we have demonstrated that stage 2 of the collision cascade is complex and can influence a detector's ultimate energy resolution. Questions persist regarding the time scale for the energy degradation from the eV to meV per quantum level. Two published calculations assume only weak perturbations of equilibrium:

we have demonstrated that is not the case. However, accepting either estimate leads to the conclusion that the gap is locally suppressed at the beginning of stage 3. The high local energy density at the end of stage 1 indicates that some phonon bottle necking is highly likely. Quasiparticle self-trapping may slow movement of the disturbed volume boundary even after the energy is degraded. Do the majority of tunneling carriers originate directly from decay of higher energy electronic excitations or from $>2\Delta$ pair breaking by phonons which escape beyond the Δ well edge? Or do those excess carriers instead merely reflect the enhanced local temperature of the cooling distrubed volume? This is an open question. Modeling stage-2 spatial propagation of both the phonons and qp in a way which explicitly incorporates energy dependence of the effective diffusion and coupling constants is highly desirable.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research. We also acknowledge several helpful conversations with Gerald Arnold of the University of Notre Dame regarding trapping and the time scale of the stage-2 expansion.

[1]See the papers and references in the following conference proceedings: (a) *Superconductive Particle Detectors,* edited by A. Barone (World Scientific, Singapore, 1988); (b) *Low Temperature Detectors for Neutrinos and Dark Matter II,* edited by L. Gonalez-Mestres and D. Perret-Gallix (Editions Frontières, Gif-sur-Yvette, 1988); (c) *Superconducting and Low Temperature Particle Detectors,* edited by G. Waysand and G. Chardin (North-Holland, Amsterdam, 1989); (d) *Low Temperature Detectors for Neutrinos and Dark Matter III,* edited by L. Brogiato, D. V. Camin, and E. Fiorini (Editions Frontières, Gif-sur-Yvette, 1990).

[2]For discussions of these works, see, for example, (a) Ref. 1; (b) D. Twerenbold, Phys. Rev. B **34**, 7748 (1986); (c) M. Kurakado and H. Mazaki, *ibid.* **22**, 168 (1980).

[3]For example, see (a) M. Kurakado and A. Matsumara, Jpn. J. Appl. Phys. **28**, L459 (1989); Appl. Phys. Lett. **57**, 1933 (1990); (b) P. Gare, R. Engelhardt, A. Peacock, D. Twerenbold, J. Lumley, and R. E. Somekh, IEEE Trans. Magn. **25**, 1351 (1989); (c) K. Ishibashi, K. Mori, K. Takeno, T. Nagae, Y. Matsumoto, S. Takada, H. Nakagawa, and H. Akoh, *ibid.* **27**, 2661 (1991); and (d) D. Van Vechten, M. N. Lovelette, C. Boyer, G. G. Fritz, M. P. Kowalski, M. Blamire, E. C. G. Kirk, and R. E. Somekh, *ibid.* **27**, 2665 (1991); **27**, 2673 (1991).

[4]See papers by D. Twerenbold; W. Rothmund and A. Zehnder; and H. Kraus, F. Pröbst, F. von Feilitzsch, and Th. Peterreins in Ref. 1.

[5]H. Kraus, F. v. Feilitzsch, J. Jochum, R. L. Mossbauer, Th. Peterreins, and F. Pröbst Phys. Lett. B **231**, 195 (1989).

[6]This form is a good description of the data in Martin J. Berger and Stephen M. Seltzer, NASA Publication NASA SP-3012 (Washington, D.C., 1964) for aluminum from 10 to 300 keV. Materials other than Al obey range-energy relations that differ by less than 3% from that given over the entire 10–300 keV initial energy range. Detailed corrections based on Z are available in J. E. Grove and R. A. Mewaldt, *21st International Cosmic Ray (IUPAP) Conference,* edited by J. Protheroe (University of Adelaide Press, Adelaide, Australia, 1990), Vol. 4, p. 398. At lower energies, only the slightly differently defined "practical ranges" have been measured, but their energy dependence is still describable as a power law with a similar exponent.

[7]Granular detectors are more affected. See Kent S. Wood and Deborah Van Vechten, Nucl. Instrum. Methods A (to be published).

[8]The majority of the parameters were taken from Roberts (B. W. Roberts, J. Phys. Chem. Ref. Data, **5**, 581 (1976); and its supplement [NBS Technical note 983 (1978)]. However the values of the Debye temperature for Hf is from S. S. Kushwaha, O.N.O. Singh, S. K. Srivastava, and N. B. Trivedi, Nuovo Cimento D **11**, 1307 (1989) and for Be is taken from T. W. Listerman and Xiao-Li Zhou, Am. J. Phys. **53**, 460 (1953). The value of 1G for $H_c(T=0)$ for Be is an estimate of Robert Soulen, Jr. (personal communication) based on his experience with the NIST temperature fixed point devices.

[9]In a superconductor, the energies of electronic excitations are $=(E^2+\Delta^2)^{1/2}$. Normally only the excitations with E small enough that the term in Δ matters are called quasiparticles. However, for convenience we herein refer to all electronic excitations as quasiparticles regardless of their energies.

[10]S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Phys. Rev. B **14**, 4854 (1976).

[11]We will discuss the time to form and significance of the quasiparticle bound-state energy level in the well in a later paper (G. Arnold, K. S. Wood, and D. Van Vechten, unpublished).

[12]N. E. Booth, Appl. Phys. Lett. **50**, 293 (1987).

[13]K. E. Gray, in *Superconductive Particle Detectors* [Ref. 1(a)].

[14]R. J. Gaitskell, D. J. Goldie, N. E. Booth, C. Patel, and G. L. Salmon, Physica (to be published).

[15]V. Narayanamurti, R. C. Dynes, P. Hu, H. Smith, and W. F. Brinkman, Phys. Rev. B **18**, 6041 (1978).

[16]Footnote 21 of C. C. Chi, M. M. T. Loy, and D. C. Cronemeyer, Phys. Rev. B **23**, 124 (1981).

[17]S. G. Han, Z. V. Vardeny, O. G. Symko, and G. Koren, IEEE Trans. Magn. **27**, 1548 (1991).

[18]U. Strom, J. G. Culbertson, P. B. Klein, and S. A. Wolf, *Optical Effects in Amorphous Semiconductors (Snowbird, Utah),* edited by P. C. Taylor and S. Bishop, AIP Conf. Proc. No. 120 (AIP, New York, 1984).

[19]M. Kurakado and H. Mazaki reported in Nucl. Instrum. Methods **185**, 141 (1981), on Monte Carlo calculation of the quasiparticle-to-phonon branching ratio and the Fano factor for tin using equilibrium distribution functions. See also W. vaan Roosbroeck, Phys. Rev. **139**, A1702 (1965) for a discussion of the meaning of the Fano factor and how else it may be calculated.

[20]That the value chosen for λ in Rothwarf-Taylor modeling matters strongly is illustrated in a paper by D. Twerenbold in Ref. 1(a). He discusses therein the sensitivity of the number of quasiparticles available to tunnel on the (single) mean free path λ used in the Rothwarf-Taylor modeling. Thanks to recombination against other nonequilibrium qp starting at $t=0$, a 10% variation in λ and a value of $\lambda=20A$ are sufficient to explain both the observed 41 eV resolution floor of tin detectors and the observed nonlinearity of the $K\alpha$-$K\beta$ line strengths.

[21]D. Van Vechten, M. N. Lovelette, C. Boyer, G. G. Fritz, M. P. Kowalski, M. Blamire, E. C. G. Kirk, and R. E. Somekh, IEEE Trans. Magn. **27**, 2665 (1991).