When "Optimal Filtering" Isn't

J. W. Fowler, B. K. Alpert, W. B. Doriese, J. Hays-Wehle, Y.-I. Joe, K. M. Morgan,
G. C. O'Neil, C. D. Reintsema, D. R. Schmidt, J. N. Ullom, and D. S. Swetz
National Institute of Standards and Technology, Boulder, Colorado 80305 USA

arXiv:1611.07856v1 [physics.data-an] 23 Nov 2016

***Abstract*—The so-called "optimal filter" analysis of a microcalorimeter's x-ray pulses is statistically optimal only if all pulses have the same shape, regardless of energy. The shapes of pulses from a nonlinear detector can and do depend on the pulse energy, however. A pulse-fitting procedure that we call "tangent filtering" accounts for the energy dependence of the shape and should therefore achieve superior energy resolution. We take a geometric view of the pulse-fitting problem and give expressions to predict how much the energy resolution stands to benefit from such a procedure. We also demonstrate the method with a case study of K-line fluorescence from several 3d transition metals. The method improves the resolution from 4.9 eV to 4.2 eV at the Cu Kα line (8.0 keV).**

I. Introduction

X-ray microcalorimeters respond electrically to the temperature change produced by the absorption of a photon. The electrical response is a brief drop in the bias current, a pulsed signal. Higher photon energies produce larger pulses. In the ideal case of a strictly linear sensor, pulses of any energy are perfectly scaled copies of each other, with the scale factor proportional to the photon energy. In the real world, pulse sizes and shapes can both depend on energy; "larger" pulses can mean pulses that have a larger peak amplitude but also last longer, if the sensor is driven past the small-signal limit.

The usual approach to the high-resolution analysis of microcalorimeter signals, however, assumes that all pulses have identical shapes. This assumption, along with others about the nature of the noise[1] leads to the procedure of optimal filtering [1]. One extracts from the signal data stream a record that contains N successive samples, with the pulse starting at some predetermined position in the record. The pulse size is then estimated by taking an inner product of the data record and the filter. The filter is thus a length-N weighting vector, one that is statistically optimal in a sense we shall see shortly.

We consider the problem of how to generalize the very successful framework of optimal filtering to improve energy resolution in cases where sensor nonlinearity is a serious obstacle. We show how to assess the expected resolution in any given data set under the new and standard analysis methods. This work considers only pulses isolated in time, but we fully expect that a deeper understanding of how to handle sensor nonlinearity will also be critical to the analysis of piled-up pulses. In short, we ask how to know when optimal filtering is, in fact, not optimal, and what can be done about it.

[1]To wit, we assume that the noise is stationary, follows a multivariate Gaussian distribution of known covariance, and is independent of the signal.

II. Optimal Filtering

First, we review optimal filtering under the usual assumption of fixed pulse shape. We will use the term pulse height to describe any estimate of the "size" of a pulse, broadly construed, and not literally the peak value of the signal. The absolute calibration of pulse heights into energy is a separate procedure, beyond the scope of this paper.

To construct the filter $\boldsymbol{f}$ that estimates pulse height, we need estimates of both the signal shape $\boldsymbol{v}$ (of length N) and the noise covariance matrix[2] $\mathbf{R}$ of size $N \times N$. Then we define the filter and the pulse height estimator, h, to be

$$\boldsymbol{f}^{\mathrm{T}} \equiv [\boldsymbol{v}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{v}]^{-1}\boldsymbol{v}^{\mathrm{T}}\mathbf{R}^{-1} \text{ and } h \equiv \boldsymbol{f}^{\mathrm{T}}\boldsymbol{d}. \qquad (1)$$

Geometrically, we can interpret h as shown in Figure 1 if we assume that the noise is white (i.e., $\mathbf{R} = \sigma^2\mathbf{I}$): the filter projects $\boldsymbol{d}$ onto the line passing through $\boldsymbol{v}$ and the origin, and h gives the coordinate along that line of the projected point.

When the noise is not white, then the projection is not the usual orthogonal projection. The $\mathbf{R}^{-1}$ factors ensure that the projection minimizes the Mahalanobis distance, rather than the Euclidean distance. Mahalanobis distance between a point and the center of some Gaussian distribution answers "how many standard deviations away?" One way to view the minimization of Mahalanobis distance is that it reduces to Euclidean distance if both data $\boldsymbol{d}$ and pulse model $\boldsymbol{v}$ are first "whitened". That is, we replace $\boldsymbol{v} \to \mathbf{W}\boldsymbol{v}$ and $\boldsymbol{d} \to \mathbf{W}\boldsymbol{d}$, where $\mathbf{W}$ is any matrix that satisfies $\mathbf{R}^{-1} = \mathbf{W}^{\mathrm{T}}\mathbf{W}$. If it is understood that the projections take place in this "noise-whitened space," then we can maintain the geometric view of the filtering operations. The filter defined here is the optimal linear estimator of pulse size, in that it is both the minimum-variance unbiased linear estimator and also the maximum-likelihood estimator, but only under the assumption of strictly constant pulse shapes.

A. Constrained optimal filters

It is often necessary to allow for a model in which pulse records are linear combinations of a few components, the usual pulse shape plus one or more additional "nuisance" components. We often find the best results in allowing for two additional components: a constant offset and a term $\mathrm{d}\boldsymbol{v}/\mathrm{d}t$, which represents a first-order correction to the pulse shape $\boldsymbol{v}$ for small variations in the pulse arrival time t. We can write the model's n linear components as the columns of the matrix

[2]I.e., R_{ij} is the expected noise covariance between samples i and j.

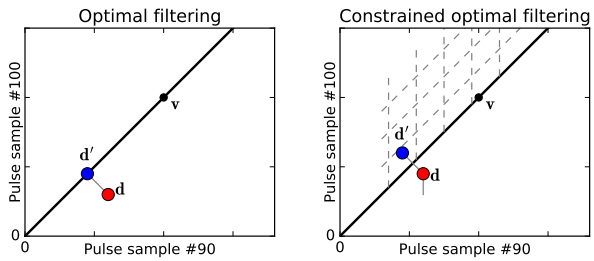

Fig. 1. A geometric view of optimal filtering without (left, Eq. 1) and with constraints (right, Eq. 2). Constrained filtering implies projection into a low-dimensional subspace (dashed coordinate system) whose coordinates are, in general, oblique. The coordinate giving distance parallel to the solid line is the pulse height, h; other coordinates correspond to nuisance terms. Only two arbitrary samples out of N are plotted, but one should imagine the points and lines to exist in an N-dimensional space.

$\mathbf{M}$; our goal is to find the vector $\boldsymbol{p}$ such that $\mathbf{M}\boldsymbol{p}$ estimates the observed signal $\boldsymbol{d}$. In this case, the optimal filter becomes [2]

$$\mathbf{F}^{\mathrm{T}} \equiv [\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}]^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}, \text{ and } \boldsymbol{p} = \mathbf{F}^{\mathrm{T}}\boldsymbol{d}. \quad (2)$$

One element of $\boldsymbol{p}$ is h, the estimated pulse height of signal $\boldsymbol{d}$. If only the pulse height is needed, then we can use a unit vector $\hat{\boldsymbol{e}}_1$ to select the appropriate column of $\mathbf{F}$ and discard the nuisance components. We call this column the constrained optimal filter $\boldsymbol{f}_c = \mathbf{F}\hat{\boldsymbol{e}}_1$, where pulse height is estimated as before: $h_{\mathrm{c}} = \boldsymbol{f}_c^{\mathrm{T}}\boldsymbol{d}$.

We can interpret $\boldsymbol{d}' \equiv \mathbf{M}\mathbf{F}^{\mathrm{T}}\boldsymbol{d}$ as a projection of the point $\boldsymbol{d}$ onto the n-dimensional subspace spanned by the columns of $\mathbf{M}$. Again, this projection will be the usual Euclidean projection only if the noise is white. For a more general noise matrix $\mathbf{R}$, we have instead that $\boldsymbol{d}'$ is the point in the subspace with the smallest Mahalanobis distance to the measured point $\boldsymbol{d}$. In general, the matrix $\mathbf{M}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}$ is not diagonal, and therefore $\boldsymbol{p}$ provides oblique coordinates in the subspace. Oblique coordinates imply higher and correlated uncertainties.

The constrained filter pulse height h_{c} is the optimal linear estimator of pulse size when the contribution of the $n-1$ unknown nuisance components must also be allowed to vary. It is optimal in that it is both the minimum-variance unbiased linear estimator and also the maximum-likelihood estimator, but again only if pulse shapes are strictly constant [3].

Assuming a known data noise covariance matrix $\mathbf{R}$, we can predict the energy resolution as a function of pulse energy. The energy uncertainty δE is a direct result of the pulse height uncertainty (let $\mathbb{E}[\cdot]$ represent expected values):

$$(\delta h)^2 = \mathbb{E}[h^2] - \mathbb{E}[h]^2 = \boldsymbol{f}_c^{\mathrm{T}}\mathbf{R}\boldsymbol{f}_c; \quad (3)$$

$$\delta E \approx \delta h \,/\, \mathbb{E}[\mathrm{d}h/\mathrm{d}E]. \quad (4)$$

To estimate $\mathrm{d}h/\mathrm{d}E$, we need to generalize the model pulse and allow $\boldsymbol{v} = \boldsymbol{v}(E)$ to be a function of energy. Then

$$\mathbb{E}[h(E)] = \boldsymbol{f}_c^{\mathrm{T}}\boldsymbol{v}(E), \quad (5)$$

$$\mathbb{E}[\mathrm{d}h/\mathrm{d}E] = \boldsymbol{f}_c^{\mathrm{T}}\mathrm{d}\boldsymbol{v}/\mathrm{d}E, \quad (6)$$

and so the expected energy resolution[3] is

$$\delta E = \sqrt{\boldsymbol{f}_c^{\mathrm{T}}\mathbf{R}\boldsymbol{f}_c} \;/\; \boldsymbol{f}_c^{\mathrm{T}}(\mathrm{d}\boldsymbol{v}/\mathrm{d}E). \quad (7)$$

The numerator is a fixed quantity because the pulse height uncertainty δh is independent of pulse energy.[4] The energy dependence appears only in the denominator, in two ways. For one, the pulse size often grows less rapidly at high energies; also, the pulse shape can change with energy, varying the angle between the vectors $\boldsymbol{f}_c$ and $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$.

III. Noise-free pulses describe a curve in $\mathbb{R}^N$

When pulse shapes change with energy, the assumptions that lead to the constrained optimal filter have failed. Will correcting these assumptions yield better energy resolution?

Consider the idealized, noise-free pulse shape as a function of energy $\boldsymbol{v}(E)$. This function describes a 1-dimensional curve in $\mathbb{R}^N$ (Fig. 2). Provided the noise is low, measured pulses will cluster in a cloud around this curve. To estimate the energy of a point $\boldsymbol{d}$ in the cloud, we should choose the "nearest" point on the curve (as before, we mean minimal Mahalanobis distance) and assign to $\boldsymbol{d}$ the energy E of that point on the curve.[5]

Many approaches to accommodate an energy-dependent pulse shape have previously been proposed. Fixsen et al. [4] have suggested the use of a discrete set of models at fixed energies; the model selected for a given pulse would be the model that produces the best fit or an appropriate linear interpolation between the two best models [5], [6]. Others have proposed use of the standard optimal filter to choose among models, followed by interpolation [7]; a fully general treatment based on differential geometry [8]; and an expansion of the signal to linear order in small changes in energy [9]. A similar linearization of the response has also been considered for use with position-sensitive detectors [10].

IV. Tangent filtering for nonlinear data

If the pulse shape $\boldsymbol{v}(E)$ is a slowly varying function of energy, then we can choose a reference point on the curve $\boldsymbol{v}_0 = \boldsymbol{v}(E_0)$ and linearize the curve in that vicinity. This will mean projecting data points $\boldsymbol{d}$ onto the line tangent to the curve at E_0. By contrast, optimal filtering in effect projects onto the secant line, the line between the origin and $\boldsymbol{v}_0$ (Fig. 2). The point projected onto the tangent line should be closer to the true $\boldsymbol{v}(E)$ curve than the same point projected onto the secant line. Whether this smaller residual yields improved energy resolution is a separate question, but it should offer clear advantages at least in the matter of identifying pathological pulse records (e.g., pulse pileup) via their rms residual.

We call this use of linearized signal tangent filtering, in that we use a filter that projects the data to the Mahalanobis-nearest point on the line tangent to $\boldsymbol{v}(E)$.

[3] This δE is a standard deviation; multiply by $\sqrt{8\ln 2}$ to get FWHM.

[4] This statement assumes that the noise does not depend on the current through the sensor, though there are physical reasons to expect some suppression of TES noise at the peak of a pulse.

[5] If the curve is parameterized by some uncalibrated quantity such as pulse height, then an additional calibration step is required, as with optimal filtering.

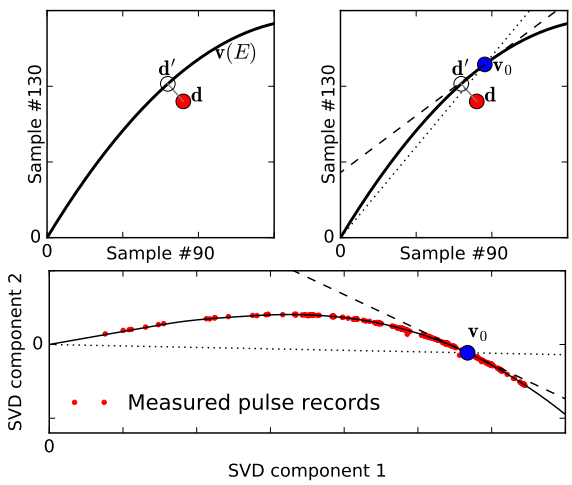

Fig. 2. When sensors are nonlinear, their response $\boldsymbol{v}(E)$ can be a curve (top left). Here, two arbitrary samples out of N are shown. To find the point, $\boldsymbol{d}'$, on the curve nearest to any given measurement, $\boldsymbol{d}$, is not the strictly linear operation of Fig. 1. Top right: given a reference pulse $\boldsymbol{v}_0$, one can approximate the nearest point to $\boldsymbol{d}$ by linear operations, by projecting $\boldsymbol{d}$ onto either the secant line (dotted) or the tangent line passing through $\boldsymbol{v}_0$ (dashed). These steps correspond, respectively, to optimal filtering and to tangent filtering. Bottom: Measurements up to 9 keV, described in Section V. The coefficients of the two leading singular vectors are shown for some 1000 pulses (dots), as well as a smooth model (solid curve) and the lines onto which data are projected by optimal filtering (dotted) and tangent filtering (dashed).

To effect tangent filtering, we operate on the data after subtracting from it the reference pulse $\boldsymbol{v}_0$. We must also replace the signal column in the n-column model matrix $\mathbf{M}$; instead of $\boldsymbol{v}_0$, the first column of $\mathbf{M}$ must be the tangent line $(\mathrm{d}\boldsymbol{v}/\mathrm{d}E)$. Call this new model matrix for the tangent-line model $\mathbf{M}_t$. Thus the tangent filter is

$$\mathbf{F}_t^{\mathrm{T}} \equiv [\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t]^{-1}\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}, \tag{8}$$

and it is applied to data $\boldsymbol{d}$ to estimate the contribution of the m components via

$$\boldsymbol{p} = \mathbf{F}_t^{\mathrm{T}}(\boldsymbol{d} - \boldsymbol{v}_0)$$

where the energy estimate is

$$\tilde{E} = \hat{\boldsymbol{e}}_1^{\mathrm{T}}\boldsymbol{p} + E_0.$$

Unlike optimal filtering, this result is already calibrated into energy (provided that the pulse curve $\boldsymbol{v}(E)$ is parameterized by a calibrated energy). This fact means that we can estimate energy resolution without needing to compute the $\mathrm{d}h/\mathrm{d}E$ conversion factor. The square of the rms resolution δE is:

$$\begin{aligned}(\delta E)^2 &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\{\mathbb{E}[\boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}] - \mathbb{E}[\boldsymbol{p}]\mathbb{E}[\boldsymbol{p}]^{\mathrm{T}}\}\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\mathbf{F}_t^{\mathrm{T}}\{\mathbb{E}[\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}] - \mathbb{E}[\boldsymbol{d}]\mathbb{E}[\boldsymbol{d}]^{\mathrm{T}}\}\mathbf{F}_t\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}\mathbf{F}_t^{\mathrm{T}}\mathbf{R}\mathbf{F}_t\hat{\boldsymbol{e}}_1 \\ &= \hat{\boldsymbol{e}}_1^{\mathrm{T}}[\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t]^{-1}\hat{\boldsymbol{e}}_1. \end{aligned} \tag{9}$$

The energy uncertainty is thus the square root of the $(1,1)$ component of the inverse of $\mathbf{A} \equiv [\mathbf{M}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{M}_t]$. We have verified Eq. 9 and Eq. 7 by applying both types of filter to simulated noise records that have the appropriate covariance.

We can get a sense of this result in the case of $n = 2$ components. Let the columns of the noise-whitened model $\mathbf{W}\mathbf{M}_t$ have magnitudes a and b and form an angle θ. Then

$$\mathbf{A} = (\mathbf{M}_t^{\mathrm{T}}\mathbf{W}^{\mathrm{T}})(\mathbf{W}\mathbf{M}_t) = \begin{pmatrix} a^2 & ab\cos\theta \\ ab\cos\theta & b^2 \end{pmatrix},$$

$$\delta E = (a\sin\theta)^{-1} = (||\mathbf{W}\,\mathrm{d}\boldsymbol{v}/\mathrm{d}E||\ \sin\theta)^{-1}.$$

Thus, the energy resolution is the inverse of the norm of the noise-whitened and -scaled pulse derivative $(\mathrm{d}\boldsymbol{v}/\mathrm{d}E)$, times a geometric factor to penalize the model insofar as it has a non-orthogonal second component. If $\mathbf{M}$ has $n \geq 3$ columns, the result is similar: $\delta E = g/a$ where $g \geq 1$ is a geometric factor that involves only the angles between noise-whitened components of the signal model. The geometric factor still penalizes the oblique coordinates, and it reduces to 1 when the coordinates are all strictly orthogonal.

The expected resolution in the cases of standard, constrained optimal filtering (Eq. 7) and tangent filtering (Eq. 9) can be compared for any given data set, so long as the following are known: the pulse size and shape as a function of energy $\boldsymbol{v}(E)$; the additional linear, nuisance components to be included in the model; and the noise covariance matrix $\mathbf{R}$. We find the resolution ratio to be a useful guide to when we can expect tangent filtering to yield an improved energy resolution.

The tangent filter certainly does not solve all problems that arise from sensor nonlinearity. Consider the special case that pulse sizes depend nonlinearly on energy, but the pulse shape is constant: $\boldsymbol{v}(E) = s(E)\boldsymbol{v}_0$. In this case, the corresponding columns of $\mathbf{M}$ and $\mathbf{M}_t$ are parallel, and both Eq. 7 and 9 reduce to the same result. Importantly, the result scales as

$$\delta E \propto (\mathrm{d}s/\mathrm{d}E)^{-1}.$$

The usual form of pulse nonlinearity is a compression, a reduction in $\mathrm{d}s/\mathrm{d}E$ as E grows; we see that this causes energy uncertainty to grow with energy. While tangent filtering is capable of recovering energy resolution lost to changes in pulse shape, it cannot help with resolution lost because of pulse size compression.

V. Case study: Kα emission up to 8 keV

Here we present a case study in which tangent filtering produces the predicted effect, a 15 % improvement in energy resolution. The data are one TES's measurements of the K-line emission of several 3d transition metals, with energies in the range of 4 keV to 8 keV. We assess energy resolution by fitting for the width of the Gaussian energy-response function in measurements of two-peaked K$\alpha_{1,2}$ line complexes.

A. Pulse modeling

Tangent filtering requires a model for the pulse size and shape versus energy, $\boldsymbol{v}(E)$. It should accurately capture $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$, at least near the energies of interest. In contrast, the usual optimal filter analysis does not require such a model—one needs only a single pulse model (typically, the average

Emission Line	Counts in peak	Energy Resolution (eV, FWHM) Predicted Optimal	Predicted Tangent	Observed Optimal	Observed Tangent
Mn Kα	4500	3.38	3.19	4.58	4.14
Fe Kα	3200	3.56	3.29	4.76	4.36
Co Kα	2500	3.78	3.42	4.53	4.17
Ni Kα	6700	4.10	3.56	4.72	3.99
Cu Kα	7800	4.36	3.69	4.92	4.20

TABLE I
ENERGY RESOLUTIONS (FWHM, IN EV), PREDICTED AND MEASURED.

over many measured pulses at a range of energies, which might well not match the pulse shape for any single energy). Therefore, we think it valuable to explain how we have constructed this model.

So that we can assign a tentative energy label to each pulse, we compute the pulse rms for each (that is, the root mean square of the measured samples, after a pre-pulse mean value is subtracted). Using the observed spectrum of pulse rms values, we find an appropriate calibration curve that converts the pulse rms into a preliminary estimate of energy.

To model $\boldsymbol{v}$ is a challenge. We need to estimate its energy derivative at each energy of interest, while the data available tend to cluster near a limited set of energies (here, the Kα and Kβ line energies). We start with the observation that the curve $\boldsymbol{v}(E)$ is in practice approximately confined to a low-dimensional subspace of the full $\mathbb{R}^N$. If we can identify a basis to span this subspace, of dimension $N_{\rm sub} \ll N$, then we need not express $\boldsymbol{v}(E)$ as N separate functions of E; a mere $N_{\rm sub}$ coefficients of that basis can approximate it instead.

Many choices of basis are possible. In the present work, we use the nine leading right singular vectors[6] from the singular-value decomposition (SVD) of a matrix whose columns are many (10^4) pulse records. We model the nine coefficients versus energy via least-squares-approximating cubic splines with ten knots between 2 keV and 9 keV, the knots more concentrated where the data are most dense (in the range of 5 keV to 7 keV). Figure 2 shows data records and the smooth model projected into the two dimensions with the highest singular values.

B. Filtering results

We apply Equations 2 and 8 to compute the constrained optimal filter $\mathbf{F}$ and the tangent filter $\mathbf{F}_t$. In the latter case, we compute three tangent filters, appropriate to the energies of the Kα lines of Mn, Co, and Cu. To reduce systematic dependence on sub-sample arrival time, we smooth the four filters with a one-pole low-pass filter ($f_{3\,\rm dB} = 12\,$kHz), after which we apply each filter to each pulse. The low-pass filtering and a gain adjustment to correct for cryogenic temperature drifts are standard steps in our best practices for achieving high energy resolution from TES microcalorimeters [11]. Finally, this TES is susceptible to cross-talk from others in the array. We cut any pulses that coincide with pulses seen in the array's other sensors (some 50 % of them), improving either method's energy resolution by nearly 1 eV.

[6]The choice of $N_{\rm sub} = 9$ is conservative. The results prove to be very insensitive to small changes in the number of basis vectors used.

The cut, filtered, and corrected data are combined into energy spectra. We use maximum-likelihood fits [12] to the known Kα line shapes [13] to extract the estimated energy resolution (Table I). The statistical uncertainty in the resolution fits is typically $\pm 0.2\,$eV for the Mn, Ni, and Cu fits, or $\pm 0.4\,$eV at the Fe and Co lines, which contain fewer photons. Besides the Gaussian resolution shown in the table, these detectors exhibit an exponential tail to low energies, a result of long-lived thermal states in the bismuth x-ray absorber.[7] The resolutions, predicted or measured, increase with energy because of pulse compression.

The observed resolution columns show that tangent filtering improves the resolution relative to the standard optimal filtering by 0.7 eV at the higher energies, and by at least 0.4 eV at all energies studied. This is a successful application of the tangent filtering method to actual TES measurements in the presence of a pulse shape that changes with energy. We attribute the remaining predicted-observed discrepancy to undetected cross-talk—some active sensors were not being recorded in this measurement—and to arrival-time systematics that arise because of the fast TES rise time (35 μs).

VI. FUTURE OUTLOOK

We have described a process to model pulse shapes and enable tangent filtering, as well as one measurement in which it provided substantially improved energy resolution: from 4.9 eV to 4.2 eV at the Cu Kα line.

Our experience suggests that improved energy resolution is not universally available, unfortunately. The angle between vectors $\boldsymbol{v}$ and $\mathrm{d}\boldsymbol{v}/\mathrm{d}E$ is generally quite small unless the TESs are operated near to reaching their normal-state resistance. We considered a variety of other data for this test; finding that the predicted resolutions improved by only a few percent, we did not attempt tangent filtering.

One question we have not explored here is one of global calibration, beyond the narrow range of a Kα complex. If we have multiple filters based on the tangent to $\boldsymbol{v}(E)$ at various E, then each will yield a different estimate of pulse energy; a procedure is needed to select or interpolate among them if an energy spectrum is to be measured over a wide range.

We plan further work to help us understand when tangent filtering improves the energy resolution, and to streamline the modeling process for routine work on large data sets produced by arrays of hundreds of detectors and over a variety of energy spectra. If such modeling succeeds, then in cases where "optimal filtering" is not actually optimal, the tangent filtering method presented here should provide a valuable improvement in the energy resolution achieved from microcalorimeter sensors.

[7]We operated several test TESs with gold absorbers in this array, which showed a dramatically smaller tail. Unfortunately, these gold-based pixels were operated far from their saturation point, so tangent filtering should not and does not appreciably improve their resolutions.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy's Office of Basic Energy Sciences, the NIST Innovations in Measurement Science program, and NASA.

REFERENCES

[1] A. E. Szymkowiak, R. L. Kelley, S. H. Moseley, and C. K. Stahle, "Signal processing for microcalorimeters," Journal of Low Temperature Physics, vol. 93, no. 3-4, pp. 281–285, Nov. 1993.

[2] B. K. Alpert, R. D. Horansky, D. Bennett, W. Doriese, J. W. Fowler, A. S. Hoover, M. Rabin, and J. N. Ullom, "Note: Operation of gamma-ray microcalorimeters at elevated count rates using filters with constraints," Review of Scientific Instruments, vol. 84, no. 5, p. 6107, May 2013.

[3] J. W. Fowler, B. K. Alpert, W. Doriese, D. A. Fischer, C. Jaye, Y. I. Joe, G. C. O'Neil, D. S. Swetz, and J. N. Ullom, "Microcalorimeter Spectroscopy at High Pulse Rates: A Multi-pulse Fitting Technique," The Astrophysical Journal Supplement Series, vol. 219, no. 2, p. 35, Aug. 2015.

[4] D. J. Fixsen, S. H. Moseley, B. Cabrera, and E. Figueroa-Feliciano, "Optimal fitting of non-linear detector pulses with nonstationary noise," Ninth International Workshop on Low Temperature Detectors, vol. 605, no. 1, pp. 339–342, Feb. 2002.

[5] D. Fixsen, S. Moseley, and B. Cabrera, "Pulse estimation in nonlinear detectors with nonstationary noise," Nuclear Instruments and Methods in Physics Research Section A, vol. 520, pp. 555–558, 2004.

[6] B. Shank, J. J. Yen, B. Cabrera, J. M. Kreikebaum, R. Moffatt, P. Redl, B. A. Young, P. L. Brink, M. Cherry, and A. Tomada, "Nonlinear optimal filter technique for analyzing energy depositions in TES sensors driven into saturation," AIP Advances, vol. 4, no. 1, p. 117106, Nov. 2014.

[7] C. Whitford, W. Bergmann Tiest, and A. Holland, "Practical considerations in optimal filtering of TES signals," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 520, no. 1, pp. 592–594, 2004.

[8] D. J. Fixsen, S. H. Moseley, T. Gerrits, A. E. Lita, and S. W. Nam, "Optimal Energy Measurement in Nonlinear Systems: An Application of Differential Geometry," Journal of Low Temperature Physics, vol. 176, no. 1-2, pp. 16–26, Mar. 2014.

[9] P. Peille, M. T. Ceballos, B. Cobo, J. Wilms, S. Bandler, S. J. Smith, T. Dauser, T. Brand, R. den Hartog, J. de Plaa, D. Barret, J. W. den Herder, L. Piro, X. Barcons, and E. Pointecouteau, " Performance assessment of different pulse reconstruction algorithms for the ATHENA X-ray Integral Field Unit," SPIE Astronomical Telescopes + Instrumentation, vol. 9905, pp. 99 055W–99 055W–14, Jul. 2016.

[10] S. J. Smith, "Implementation of complex signal-processing algorithms for position-sensitive microcalorimeters," Nuclear Instruments and Methods in Physics Research Section A, vol. 602, no. 2, pp. 537–544, Apr. 2009.

[11] J. W. Fowler, B. K. Alpert, W. Doriese, Y. I. Joe, G. C. O'Neil, J. N. Ullom, and D. S. Swetz, "The Practice of Pulse Processing," Journal of Low Temperature Physics, vol. 184, no. 1, pp. 374–381, Jul. 2016.

[12] J. W. Fowler, "Maximum-Likelihood Fits to Histograms for Improved Parameter Estimation," Journal of Low Temperature Physics, vol. 176, no. 3, pp. 414–420, Aug. 2014.

[13] G. Hölzer, M. Fritsch, M. Deutsch, and J. Härtwig, "K$\alpha_{1,2}$ and K$\beta_{1,3}$ x-ray emission lines of the 3d transition metals," Physical Review A, vol. 56, no. 6, pp. 4554–4568, 1997.