Energy Calibration of Nonlinear Microcalorimeters with Uncertainty Estimates from Gaussian Process Regression

J. W. Fowler[1,2] ⬤ · B. K. Alpert[1] · G. C. O'Neil[1] · D. S. Swetz[1] · J. N. Ullom[1,2]

Abstract

The nonlinear energy response of cryogenic microcalorimeters is usually corrected through an empirical calibration. X-ray or gamma-ray emission lines of known shape and energy anchor a smooth function that generalizes the calibration data and converts detector measurements to energies. We argue that this function should be an approximating spline. The theory of Gaussian process regression makes a case for this functional form. It also provides an important benefit previously absent from our calibration method: a quantitative uncertainty estimate for the calibrated energies, with lower uncertainty near the best-constrained calibration points.

Keywords Microcalorimeters · Detector calibration · Superconducting detectors

1 Introduction

Cryogenic microcalorimeters have been used to measure x-ray and gamma-ray emission in many diverse settings. Considering only existing x-ray spectrometers based on the transition-edge sensor (TES), these applications include [1] synchrotron beamlines, astrophysical telescopes, exotic atom measurements at particle accelerators, computed tomography with spectral discrimination, and even metrological study of fluorescence energies. One requirement these applications share is accurate energy estimation for each photon detected, a difficult challenge given that the measurements are made with sensors whose energy response is inherently nonlinear.

Users of such devices generally employ an empirically estimated *calibration curve*, a function that converts each measured pulse height to a photon energy. The

---

✉ J. W. Fowler
   joe.fowler@nist.gov

1   National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA

2   Department of Physics, University of Colorado, Boulder, CO 80309, USA

pulse height is typically estimated by statistically optimal filtering, corrected for confounding effects such as slow drifts in system gain or for a bias that depends on the relative arrival time of a photon and the system's sampling clock [2]. We will call this optimal, corrected value simply the *pulse height* (PH, or p in equations). The calibration curve is "anchored" by a small number of spectral features. These features must have an absolute energy that is known a priori, and they must be detected with high enough intensity that their representative pulse height can be estimated from the measured spectrum with small uncertainties. Such features might be gamma rays, x-ray fluorescence line emission, or light emitted by a calibrated x-ray monochromator.

To create a calibration curve from a set of anchor points, one must make several decisions. Should the curve interpolate or approximate the data? Should it be a polynomial, a spline, or another functional form? How should parameters such as the degree of the polynomial or the number and locations of the spline's knots be chosen? Should the smoothing function relate energy to PH directly or indirectly?

We think the answer to the first question is clear: it is best to have a calibration curve approximate the data. Anchor points have uncertainty in both the PH and energy, and only an approximating curve can fully account for the uncertainties. With polynomial fits, approximation is accomplished by weighted least-squares fits to a polynomial of degree lower than the number of anchor points. Though splines normally interpolate their data, several methods to approximate with splines exist; we have previously found best results with a cubic smoothing spline [3].

Our most important result is the derivation of the *uncertainty* on the calibration function from the theory of Gaussian process regression (GPR). It quantifies our intuition that the energy scale is more uncertain at energies that are further from the anchor points, or near to anchor points that themselves have higher uncertainty. GPR also justifies the smoothing spline form.

The procedure described here has been tested only with data from TES microcalorimeters with Mo–Cu bilayers and designed with normal-metal banks and bars. Such sensors have nonlinear energy responses that are difficult or impossible to model a priori, and empirical calibrations are necessary. We expect that the GPR procedure should generalize to other TES designs or other sensors, but without testing other cases, we cannot state when it will be needed.

2 Choosing a Calibration Space

The calibration curve is a function that estimates the photon energy from a PH value p. It can employ a spline directly or indirectly. Indirect use means the spline models some $y(E, p)$ as a function of some $x(p)$, so long as $x(\cdot)$ is a monotone function of its input and $y(E, p)$ is a function for which $y_0 = y(E, p)$ can be easily solved for E given y_0 and p. Many functions x and y, or *calibration spaces* are possible. Here we consider p and $\log p$ for $x(p)$; and E, p/E, E/p, and $\log E$ for $y(E, p)$.

The x and y functions can be picked by the principle that the spline function should "do the least work." For example, in the case of TES calibration from 4 to 10 keV for a study of rare-earth L lines [3], we found it best to spline $y(E, p) = p/E$,

Table 1 Median energy error across all TESs when the linear interpolation of two points is used at an intermediate point, for seven calibration spaces

Calibration space		Median linear-interpolation error (eV)			
$x(p)$	$y(E, p)$	Ti/Cr/Fe	V/Mn/Co	Cr/Fe/Ni	Mn/Co/Cu
p	E	16.8	18.8	20.2	22.5
p	p/E	0.5	0.5	0.7	1.3
p	E/p	2.1	2.6	3.0	4.0
p	$\log(p/E)$	1.3	1.6	1.9	2.7
$\log p$	p/E	7.9	8.7	9.3	10.5
$\log p$	E/p	9.5	10.7	11.6	13.3
$\log p$	$\log E$	8.7	9.7	10.5	11.9

The column "Ti/Cr/Fe" uses linear interpolation of the Kα lines of Ti and Fe to estimate the Cr Kα energy, and similarly for the last three columns. Each interpolation spans approximately 1 keV. The intermediate elements have Kα energies of 5.4 keV (Cr), 5.9 keV (Mn), 6.4 keV (Fe), and 6.9 keV (Co). The simplest calibration space (row 1, E vs. p) introduces large errors. The smallest errors arise when gain p/E vs. p is taken to be linear (row 2). The use of $x(p) = \sqrt{p}$ (not listed) gives results intermediate between p and $\log p$. In any calibration space, the linear-interpolation error is larger at higher photon energy for these data (from a study of lanthanide metal emission [3])

which we called the sensor *gain*, as a function of $x(p) = p$. Consider the Kα lines of Mn, Co, and Cu at approximately 6 keV, 7 keV, and 8 keV. We asked how far in error the intermediate (Co Kα) energy estimate would be if the curve were the linear interpolation of the Mn and Cu Kα peaks for each possible choice of y and x functions. Only p/E vs. p was correct to within 1.3 eV for the median detector; linear interpolations of other calibration spaces yielded errors up to 23 eV (Table 1). It is possible that other calibration spaces would be preferred for other measurements, because the least-curvature space potentially depends on any of: the detector material and shape, its bias voltage, cryogenic bath temperature, and the energy range of interest. The three-point test of linear-interpolation error offers a simple way to choose. A second benefit of choosing $x = p$ and $y = p/E$ (or $y = E/p$ or $y = \log(p/E)$) is that any finite, positive value of y when $x = p = 0$ will guarantee that the calibration curve yields the expected energy $E = 0$ for pulses of size $p = 0$.

3 Smoothing Splines for Approximation

In previous work [3, 4], we argued that the smoothing spline is the best way to generalize the calibration x–y relationship. It does not interpolate the anchor points exactly but strikes a compromise between fidelity to the data and minimal curvature. We assume that the more a spline has to curve, the poorer a model it is for calibration—particularly if the y and x functions are chosen to require minimal curvature, as proposed in the previous section. Of all twice-differentiable functions, a

cubic smoothing spline is the one [5] that minimizes the penalty functional ("cost function")

$$C[h] \equiv \sum_{i=1}^{n} \left(\frac{h(x_i) - y_i}{\sigma_i} \right)^2 + \lambda \int_a^b |h''(x)|^2 \, dx \qquad (1)$$

where $x_1 \ldots x_n$ and $y_1 \ldots y_n$ are the anchor points and σ_i is the uncertainty on y_i. The first term in C is the usual χ^2 statistic for disagreement between data and model, as we expect $h(x_i) \approx y_i \pm \sigma_i$. The integral is the model's curvature over the range $a = \min(x_i)$ to $b = \max(x_i)$. A regularization parameter λ controls the balance between data fidelity and minimization of curvature. For $\lambda = 0$, curvature is not penalized, and minimization of C produces an interpolating spline. In the limit $\lambda \to \infty$, curvature is forbidden, and h will be a line: the line that minimizes the (uncertainty-weighted) sum of squared error between data and model.[1] For finite λ, a cubic[2] smoothing spline results. It is a cubic spline function $h(x)$ with n knots at x_i and natural boundary conditions—that is, $h''(x) = 0$ at the lowest and highest values of x_i.

This method raises two important questions. First, what curvature penalty λ is appropriate? For Gaussian errors, the first term in Eq. 1 has expected value $E[\chi^2] = n$. A λ is reasonable if it yields $\chi^2 \approx n$ when cost C is minimized, but we could use a more principled approach. Second, what is the calibration uncertainty on the minimum-cost curve? Gaussian process regression answers both questions.

4 Calibration Curves as a Gaussian Process

A Gaussian process (GP) model describes a distribution over functions. It posits that the distribution of function *values* at any finite set of points **x** in the domain is a multivariate Gaussian, characterized by its mean and covariance. It generalizes the mean from a vector to a function $m(x)$ and the covariance from a matrix to a function of two variables $k(x, x')$. The initial distribution is consistent with a broad class of models, not informed by the observed data **y**. After observations,[3] the mean and covariance are made to be consistent with the data. The mean now estimates the calibration function $h(x)$ we want to learn; the covariance characterizes its uncertainty. This formulation is a Bayesian framework, with the a priori distribution modified by the observed data to yield an a posteriori distribution. The posterior distribution enables computation of expected values and covariances for any set of points $\mathbf{x}_\star$. GPR means the refinement, or regression, of a GP model given the data [7, 8].

The GPR analysis of h would immediately solve the second difficulty we have with calibration splines, because GPR provides both a model function—the expected

---

[1] When the data can be exactly interpolated by a line, that line is found for any value of λ.

[2] Defining curvature as the integral of the kth derivative squared yields [6] splines of degree $(2k - 1)$.

[3] Estimates of the uncertainty on the measurements are also required, for which we use the simplest possible model: that the noise is independent and Gaussian-distributed with mean zero and variance σ_i^2.

value of $h(x)$—and an uncertainty measure—the expected covariance of h at any two points (x, x'). Choosing the mean and covariance functions m and k to specify the model space might at first seem a more open-ended problem than the question that ended the previous section: how much to penalize curvature.

Remarkably, *there is a perfectly plausible GP model space that yields smoothing splines as the expected GPR function values*: those functions that are the integral of a continuous random walk (that is, a once-integrated Wiener process) over the interval bounded by the x_i values plus a line over the entire real domain [6]. Integrating a Wiener process once allows the expected slope to differ at the two ends of the measured interval. Such a GP model has expected value and covariance[4]

$$m(x) \;\; = \beta_0 + \beta_1 x \quad \text{and} \tag{2}$$

$$k(x, x') \;\; = \sigma_f^2 [v^3/3 + v^2 |x - x'|/2] \tag{3}$$

where $v \equiv \min(x, x')$. The two parameters (β_0, β_1) of the mean are assumed to have a "diffuse" or uninformative Bayesian prior, so that all possible lines are equally probable. The parameter σ_f^2 that scales the covariance function controls the expected amount of curvature; a larger σ_f^2 corresponds to higher curvature.

This GP space is a reasonable way of modeling functions that are linear on either side of the measured interval (potentially with different slopes), and as little structure as possible inside it. Given that the space is reasonable and yields the same, convenient smoothing spline functions we have already shown to work quite well, it is the model space we use for calibration curves.

What curvature scale σ_f^2 should be used for the Wiener process (random walk) component in Eq. 3? This is equivalent to the question, what curvature penalty λ should be used in the smoothing spline optimization of Eq. 1? Rasmussen and Williams [8] (R &W) show that our regularization parameter λ (Eq. 1) and the curvature scale σ_f^2 (Eq. 3) are related by $\lambda \sigma_f^2 = 1$. They also give an expression for the marginal likelihood, i.e., the likelihood after integrating (marginalizing) over the values of the spline function at its knots. This marginal likelihood is a function of σ_f^2 and appears below as Eq. 8 (or as R &W Equation 2.45). The value of σ_f^2 that maximizes the marginal likelihood is the value most consistent with the data.

Equations 2 and 3 yield a GP model that R &W call GPR with a basis set. For measurement points $\mathbf{x}$ and sample points $\mathbf{x}_\star$, we have measurements $\mathbf{y}$ and sample point values $\mathbf{h}_\star$ jointly distributed as the multivariate Gaussian

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{h}_\star \end{bmatrix} \sim \mathcal{N} \left(\beta_0 + \beta_1 \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_\star \end{bmatrix}, \begin{bmatrix} \mathbf{K_y} & \mathbf{K_\star} \\ \mathbf{K_\star^T} & \mathbf{K_{\star\star}} \end{bmatrix} \right), \tag{4}$$

prior to marginalizing over the observed measurements $\mathbf{y}$, where $\mathbf{K}$ is the GP covariance matrix with $K_{ij} = k(x_i, x_j)$; $\mathbf{K}_y = \mathrm{diag}(\boldsymbol{\sigma}^2) + \mathbf{K}$ is the measurement noise plus

---

[4] Here the covariance is simplified by assuming the domain is transformed to $[\min x_i, \max x_i] = [0, 1]$.

GP covariance; $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_n^2]^T$; $(K_\star)_{ij} = k(x_i, x_{\star j})$; and $(K_{\star\star})_{ij} = k(x_{\star i}, x_{\star j})$. The marginalized, predictive distribution is $\mathbf{h}_\star | (\mathbf{x}, \mathbf{y}, \mathbf{x}_\star) \sim \mathcal{N}\left(\overline{\mathbf{h}}_\star, \text{cov}(\mathbf{h}_\star)\right)$ where

$$\overline{\mathbf{h}}_\star \equiv \mathbb{E}[\mathbf{h}_\star | \mathbf{x}, \mathbf{y}, \mathbf{x}_\star] \quad = \mathbf{m}(\mathbf{x}_\star) + \mathbf{K}_\star^T \mathbf{K}_y^{-1} [\mathbf{y} - \mathbf{m}(\mathbf{x})] \tag{5}$$

$$\text{cov}(\mathbf{h}_\star) \quad = \mathbf{K}_{\star\star} - \mathbf{K}_\star^T \mathbf{K}_y^{-1} \mathbf{K}_\star \tag{6}$$

with $\mathbf{m}(\mathbf{x}) = \beta_0 + \beta_1 \mathbf{x}$.

In Eqs. 2 and 3, the parameters $\beta = \langle \beta_0, \beta_1 \rangle^T$ and σ_f^2 were given, but we want to choose them to maximize the likelihood of the measurements $\mathbf{y}$. Equation 4 shows that the nonzero mean of the GP is linear in the parameters β. Suppose a regression will be done with n calibration anchor points and two basis functions: the pair $h_0(x) = 1$ and $h_1(x) = x$. Let $\mathbf{H}$ be the $2 \times n$ matrix whose rows are the basis functions at the n locations $\mathbf{x}$. For $\beta \sim \mathcal{N}(\mathbf{b}, \mathbf{B})$, minimization of the expected error with respect to β yields $\overline{\beta} = \left(\mathbf{H}\mathbf{K}_y^{-1}\mathbf{H}^T + \mathbf{B}^{-1}\right)^{-1} \left(\mathbf{H}\mathbf{K}_y^{-1}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b}\right)$. An uninformative Bayesian prior implies $\mathbf{B}^{-1} \rightarrow \mathbf{0}$, for which $\mathbf{b}$ becomes irrelevant, and

$$\overline{\beta} = \left(\mathbf{H}\mathbf{K}_y^{-1}\mathbf{H}^T\right)^{-1} \mathbf{H}\mathbf{K}_y^{-1}\mathbf{y}. \tag{7}$$

Wahba [6] has shown that the expected function values $\overline{\mathbf{h}}_\star$ (Eq. 5), with $\beta = \overline{\beta}$, is a cubic spline of $\mathbf{x}_\star$ with knots at the measurement locations $\mathbf{x}$ and with natural boundary conditions. Therefore, we can use a shortcut. Instead of evaluating the function for every value of $x_\star$ of interest, we can evaluate it only n times—by choosing sample points $\mathbf{x}_\star$ to coincide with the knots $\mathbf{x}$. The approximating function we seek is the unique cubic spline with natural boundary conditions that interpolates these n predictions (which are near to but not exactly the values $\mathbf{y}$).

To use the above equations, we need the factor σ_f^2 that scales the covariance function $k(x, x')$ and thus every entry in $\mathbf{K}$, $\mathbf{K}_\star$, and $\mathbf{K}_{\star\star}$. We set σ_f^2 to the value that maximizes the marginal likelihood (the Bayesian probability of measuring $\mathbf{y}$), obtained by a somewhat elaborate calculation described in R &W:

$$2 \log P\left(\mathbf{y} | \mathbf{x}, \sigma_f^2\right) = -\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y} + \mathbf{y}^T \mathbf{C}\mathbf{y} - \log |\mathbf{K}_y| - \log |\mathbf{A}| - (n-2) \log 2\pi \tag{8}$$

where $\mathbf{A} \equiv \mathbf{H}\mathbf{K}_y^{-1}\mathbf{H}^T$ and $\mathbf{C} \equiv \mathbf{K}_y^{-1}\mathbf{H}^T \mathbf{A}^{-1} \mathbf{H}\mathbf{K}_y^{-1}$.

5 Example Curves

Figure 1 shows the results of one such calibration procedure, with a cubic smoothing spline of gain $g = p/E$ vs. p. The calibration curve is shown both directly and with a linear trend subtracted to emphasize the difference between the trend and the complete calibration curve. The figure also illustrates the varying GPR-estimated uncertainty associated with anchor points not uniformly spaced and with unequal
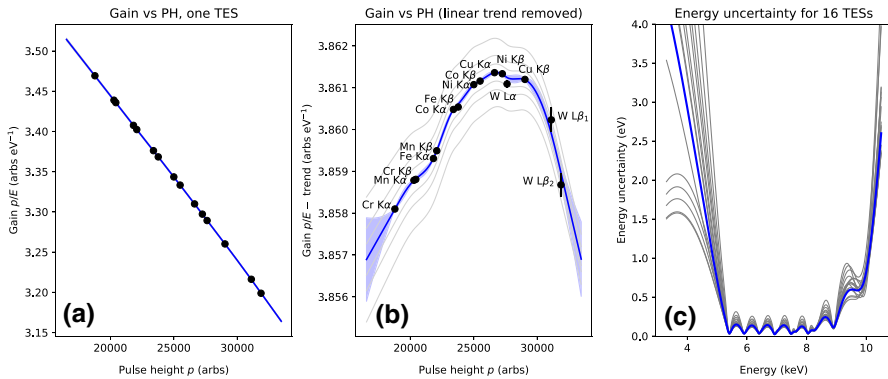
Fig. 1 **a** Example calibration curve (gain vs. PH) for one representative TES. **b** Same data as **a**, except here the linear trend of $-2.07 \times 10^{-5} \, \text{eV}^{-1}$ is subtracted from gain, to highlight departures from the trend. Labeled points (•) with 1σ error bars are the 15 anchor points used. The solid curve is the smoothing spline (Eq. 5), and the shaded band represents the $\pm 1\sigma$ calibration uncertainty (square root of Eq. 6). Six thin gray curves are placed at ± 0.5 eV, ± 1 eV, and ± 2 eV about the best calibration to indicate the energy scale. **c** The calibration curves' energy uncertainty as a function of energy for 16 representative TESs (the heavier line corresponds to the same TES featured in **a** and **b**). Anchor points range from 5.4 to 10 keV. The calibration uncertainty is lower closest to the anchor points, particularly those measured with low uncertainties. The five sensors with smaller uncertainties below 4 keV are those that measured enough Si Kα emission at 1.7 keV to have an additional anchor point at that energy. (Color figure online.)

uncertainties. Through most of the energy range populated by anchor points, 5.4–9 keV, the calibration uncertainty is well below 0.5 eV, as are the small-scale features in the curve. As expected, the uncertainty is smallest at the well-measured anchor points and grows with distance from the nearest one. Changing the variance σ_f^2 by a factor of $2^{\pm 1}$ relative to its maximum-likelihood value moves energy estimates by no more than ± 0.1 eV in most cases.

6 Conclusion

We have considered the construction of energy–calibration curves for nonlinear TES microcalorimeters in the framework of Gaussian processes. With the once-integrated Wiener process as our specific GP model, the predictions follow a cubic smoothing spline with natural boundary conditions. We have previously shown that such an approximating spline is an excellent match to TES calibration [3], but the GPR framework also permits computation of the calibration uncertainties.

We can briefly summarize the calibration method:

1. Using triples of calibration points, learn which calibration space has the least curvature for the current data (Sect. 2).
2. Find the maximum-likelihood GP variance, σ_f^2 (maximize Eq. 8).
3. Compute the predictions $\overline{h}(\mathbf{x})$ for each anchor point (Eq. 5).
4. Calibration $h(x)$ is the cubic smoothing spline that interpolates those predictions.

Item 1 should be done once for the given measurement and detector array; the other steps are repeated once for each calibration (say, once per sensor per day). We plan further investigation into whether it is better to use a separate σ_f^2 for each sensor or a single, universal value. We intend to use the calibration method based on GPR, and the energy uncertainty it generates, for future TES calibrations.

References

1. W. Doriese et al., A practical superconducting-microcalorimeter X-ray spectrometer for beamline and laboratory science. Rev. Sci. Instrum. **88**, 053108 (2017)
2. J.W. Fowler, B.K. Alpert, W. Doriese, Y.-I. Joe, G. O'Neil, J. Ullom, D. Swetz, The practice of pulse processing. J. Low Temp. Phys. **184**, 374 (2016)
3. J.W. Fowler et al., Absolute energies and emission line shapes of the L x-ray transitions of lanthanide metals. Metrologia **58**, 015016 (2021)
4. J.W. Fowler et al., A reassessment of absolute energies of the x-ray L lines of lanthanide metals. Metrologia **54**, 494–511 (2017)
5. P.J. Green, B.W. Silverman, *Nonparametric Regression and Generalized Linear Models* (Chapman and Hall, London, 1994)
6. G. Wahba, Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. R. Stat. Soc.: B Methodol. **40**, 364–372 (1978)
7. K.D. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge, 2012)
8. C.E. Rasmussen, K.I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.