Fluorescent X-ray Scan Image Quality Prediction

Peter B. Weichman¹ · Eugene M. Lavelly¹

Received: 6 June 2019 / Accepted: 2 October 2019 / Published online: 15 November 2019
© The Author(s) 2019

Abstract

This paper summarizes approaches to image quality prediction in support of an effort under the IARPA RAVEN program to demonstrate a non-destructive, tabletop X-ray microscope for high-resolution 3D imaging of integrated circuits (ICs). The fluorescent X-rays are generated by scanning an electron beam along an appropriately patterned target layer placed in front of the sample and are then detected after passing through the sample by a high-resolution (in both solid angle and energy) backside sensor array. The images are created by way of a model-based tomographic inversion algorithm, with image resolution depending critically on the electron beam scan density and diversity of sample orientations. We derive image quality metrics that quantify the image point spread function and noise sensitivity for any proposed experiment design. Application of these metrics will guide final system design when physical data are not yet available.

Keywords Image quality · X-ray tomography · Scanning electron microscope (SEM) · Transition edge sensor (TES) · Integrated circuit (IC)

1 Introduction

Image quality (IQ) assessment is an important element of system design, optimization, and quality control [1, 2]. A complete assessment evaluates the entire imaging chain including both data acquisition and image reconstruction stages. *Retrospective* IQ measurement is based on careful analysis of images taken in a series of laboratory measurements, often using specialized phantoms. In contrast, *prospective* image quality prediction is based on a high-quality end-to-end simulation of the system data product, and subsequent image reconstruction algorithm applied to that data. A key application is to guide system design in advance of hardware construction when physical data are not yet available. Similarly, it allows one to tune or

customize an existing system to a particular set of specialized tasks under conditions where the retrospective approach is time-consuming or infeasible for other reasons (e.g., target sensitivity to, or potential degradation under, multiple measurements).

Given an accurate system forward model, IQ assessment is relatively straightforward in cases of “direct” image reconstruction algorithms, in which a relatively simple (though perhaps numerically intensive) forward algorithm is applied to the data to obtain an image. An example is back projection image formation, which is also linear in the data. Various Fourier transform methods (e.g., SAR or range-Doppler processing) are special cases. Construction of various IQ metrics, such as point spread functions, are equally direct.

Less straightforward are imaging algorithms that rely on model-based iterative reconstruction (MBIR). The resulting algorithm may be highly nonlinear in the data, and IQ metrics may therefore vary strongly with position, noise levels, source and detector blur, etc. This certainly increases the complexity of, and necessity for, IQ assessment. Here, we develop an IQ assessment approach for the fluorescent X-ray tomography system (Fig. 1). We define and evaluate a set of rigorous IQ metrics informed by the MBIR based on the Poisson statistics of photon counting. Specific forward model elements for this system that must be accurately modeled include focal spot blur (finite-size X-ray source

Distribution Statement A. Approved for public release: distribution unlimited.

✉ Peter B. Weichman
peter.weichman@baesystems.com

Eugene M. Lavelly
eugene.lavelly@baesystems.com

¹ Fast Labs, BAE Systems, Electronic Systems, 600 District Avenue, Burlington, MA 01803, USA

Fig. 1 Top: Illustrated is the microscope interior chamber, highlighting the electron column, electron detectors (SE, BSE), and the TES-based [3] transmission X-ray detector assembly. The target/sample assembly is mounted in the sample holder and positioned for imaging with a translation and rotation stage. The detector package is mounted in a bellows inserted into a “receiver tube” for high proximity to the X-ray source. The instrument is a customized adaptation of Orsay Physics’ “NanoSpace” instrument, performed collaboratively with Orsay Physics. Bottom: Target/sample assembly and stimulated signals. The tomographic data collection uses fluorescent X-rays induced in a metallic thin film deposited onto the prepared backside of the sample. To obtain desired resolution, the sample must be decapsulated and the silicon backside thinned to within a few hundred nanometers of the semiconductor device layer prior to the deposition step

region in the target), cross-talk between superconducting detector pixels when recording change in conductivity due to a single photon capture, and noise correlations.

This paper is intended to provide a detailed theoretical basis for a general set of IQ metrics. Specific illustrative applications to the RAVEN IC tomography problem will be presented separately.

2 System Forward Model

In the X-ray microscope system (Fig. 1), the collected fluorescent photon count data forms the basis for a tomographic inversion of the sample structure. We begin by accounting for the physics and geometry of the measurement, assuming ideal sample stage (translation and rotation) operation, perfect knowledge of the electron beam location, and ideal photon detector behavior. Later, we will include effects of sample stage and electron beam uncertainty, as well as non-ideal detector effects such as pixel cross-talk and noise properties.

2.1 Measurement Physics

In what follows, we choose a 3D coordinate system that is fixed in the sample. Of course, it is generally the sample that is translated and rotated while the electron beam and detector apparatus remain fixed. However, it is the detailed sample structure that is the desired end-product of the tomographic inversion, and these take the form of fixed functions of position within the sample volume. In this frame, the relative positions and orientations of the source, receiver, and electron beam must be carefully tracked for each measurement.

For an idealized point source $\mathbf{x}_S$ and point receiver $\mathbf{x}_R$, the mean count rate for a chosen fluorescent photon with sharp energy E is modeled in the form

$$n(\mathbf{x}_S, \mathbf{x}_R; E) = n_0(\mathbf{x}_S, E) \frac{e^{-M(\mathbf{x}_S, \mathbf{x}_R; E)}}{4\pi |\mathbf{x}_R - \mathbf{x}_S|^2} + n_B(\mathbf{x}_S, \mathbf{x}_R; E) \Delta E \quad (1)$$

where n_0 is the raw fluorescent photon production rate, n_B is the background rate (due to bremsstrahlung, multiple

scattering, etc.) of continuum photons deposited in the energy bin ΔE about E . The first term represents the photon intensity derived from the physical optics approximate solution to the 3D wave equation, including geometric spreading and energy dissipation. For the latter, fluorescent photons are absorbed (or scattered to lower energy) according to the line integral (valid in the weak absorption limit characteristic of higher energy X-rays)

$$M(\mathbf{x}_S, \mathbf{x}_R; E) = \int_0^{|\mathbf{x}_R - \mathbf{x}_S|} \mu[\mathbf{x}(s); E] ds \quad (2)$$

in which

$$\begin{aligned} \mathbf{x}(s) &= \mathbf{x}_S + s\hat{\mathbf{e}}_{SR} \\ \hat{\mathbf{e}}_{SR} &= \frac{\mathbf{x}_R - \mathbf{x}_S}{|\mathbf{x}_R - \mathbf{x}_S|} \end{aligned} \quad (3)$$

defines the ray, parameterized by physical distance $0 \leq s \leq |\mathbf{x}_R - \mathbf{x}_S|$, connecting the two endpoints. The function $\mu(\mathbf{x}, E)$ is the local absorption rate as a function of 3D location $\mathbf{x}$, commonly modeled as a linear superposition

$$\mu(\mathbf{x}) = \sum_A \mu_A(E) n_A(\mathbf{x}), \quad (4)$$

in which the index A ranges over atomic elements (Au, Al, Si, Cu, etc.) and $n_A(\mathbf{x})$ is the element number density.

For specified electron beam entry center point $\mathbf{x}_S^0$, we write the position-dependent fluorescent production rate in the form

$$n(\mathbf{x}_S, E) = \bar{n}_0(\mathbf{x}_S^0) \Phi_S(\mathbf{x}_S - \mathbf{x}_S^0; E, \mathbf{x}_S^0, \hat{\mathbf{e}}_S^0) \quad (5)$$

in which the source distribution function Φ_S is normalized to unit integral in the first argument, and is now explicitly parameterized as well by the electron beam energy E (through fluorescent production efficiency, electron multiple scattering, etc.), the beam center point $\mathbf{x}_S^0$, and the incident direction $\hat{\mathbf{e}}_S^0$ of that beam. Thus,

$$\bar{n}_0(\mathbf{x}_S^0) = \int d\mathbf{x}_S n_0(\mathbf{x}_S, E) \quad (6)$$

represents the total integrated production rate, in general also depending on energy and incident beam direction, but this will normally be suppressed from the notation.

We now use Φ_S to define an average of any function $F(\mathbf{x}_S, \mathbf{x}_R)$ of the source and receiver points by

$$\begin{aligned} \langle F(\mathbf{x}_S, \mathbf{x}_R) \rangle_0 &\equiv \bar{F}(\mathbf{x}_S^0, \mathbf{x}_R^0; E, \hat{\mathbf{e}}_S) \\ &= \int d\mathbf{x}_S \Phi_S(\mathbf{x}_S - \mathbf{x}_S^0; E, \mathbf{x}_S^0, \hat{\mathbf{e}}_S^0) \\ &\quad \times \int_{A_R} \frac{d\mathbf{a}_R \cdot \hat{\mathbf{e}}_{SR}^0}{\mathbf{A}_R \cdot \hat{\mathbf{e}}_{SR}^0} F(\mathbf{x}_S, \mathbf{x}_R), \end{aligned} \quad (7)$$

which includes separate averages over the 3D source region volume and over the 2D receiver pixel area A_R . The result

is defined by the entry center point $\mathbf{x}_S^0$ of the electron beam on the target surface, the receiver center coordinate $\mathbf{x}_R^0$, and the corresponding mean incident X-ray flux direction $\hat{\mathbf{e}}_{SR}^0 = \frac{\mathbf{x}_R^0 - \mathbf{x}_S^0}{|\mathbf{x}_R^0 - \mathbf{x}_S^0|}$. The receiver area integral is normalized by the total projected area $\mathbf{A}_R \cdot \hat{\mathbf{e}}_{SR}^0$ along the incident X-ray flux direction. More generally, one could also incorporate a receiver sensitivity profile that is not uniform over the pixel.

With the above definitions, for finite-size source and receiver, the actual mean measured count rate is expressed in the form

$$\begin{aligned} \bar{n}_{\text{meas}}(\mathbf{x}_S^0, \mathbf{x}_R^0; E) &= \int d\mathbf{x}_S \int_{A_R} d\mathbf{a} \cdot \hat{\mathbf{e}}_{SR}^0 n(\mathbf{x}_S, \mathbf{x}_R; E) \\ &= \frac{\Delta\Omega_{SR}}{4\pi} \bar{n}_0(\mathbf{x}_S^0) \langle e^{-M(\mathbf{x}_S, \mathbf{x}_R; E)} \rangle_0 \\ &\quad + \bar{n}_B(\mathbf{x}_S^0, \mathbf{x}_R^0; E) \Delta E \end{aligned} \quad (8)$$

in which $\bar{n}_B$ is the integrated background rate, and

$$\Delta\Omega_{SR} = \frac{\mathbf{A}_R \cdot \hat{\mathbf{e}}_{SR}^0}{|\mathbf{x}_S^0 - \mathbf{x}_R^0|^2} \quad (9)$$

is the solid angle subtended by the receiver pixel—in the denominator, we have simply substituted $\mathbf{x}_S - \mathbf{x}_R \rightarrow \mathbf{x}_S^0 - \mathbf{x}_R^0$ under the reasonable assumption that source region and pixel diameter are very small compared to their separation. The background term is not resolved in detail here since it is more difficult to model quantitatively and is assumed to be a smooth function of E , which can be fit and subtracted directly from the data. Although in principle present, any residual dependence of the background on μ is neglected.

The function Φ_S encompasses the entire fluorescent X-ray creation process, including incident electron beam shape and energy spectrum, fluorescing layer thickness, and orientation. All of these need to be included in the system model. For a horizontally homogeneous target, the extra dependence of Φ_S on $\mathbf{x}_S^0$ drops out (with only the difference variable $\mathbf{x}_S - \mathbf{x}_S^0$ surviving). However, target inhomogeneity will in general be present (e.g., through varying deposited film thickness).

The source region may be crudely thought of as a cylinder with axis oriented along $\hat{\mathbf{e}}_S^0$ and diameter determined by electron beam scattering characteristics. More generally, Φ_S would either need to be determined via prior calibration measurements in the absence of the sample, or modeled through Monte Carlo simulations of the electron beam, with given energy and geometry, interacting with the target with given atomic composition and geometry [4]. Clearly, detailed calibration measurements would be most desirable, but these are unlikely to be available given that the target film is most likely directly applied to the sample in our setup (Fig. 1). In absence of this, an effort should be made to produce as uniform a film as possible. With sufficient measurement diversity, it turns out that although it

is essential that the electron beam position be tracked to high precision, the source intensity (hence film thickness) may to a certain extent be included as part of the inversion. Thus, given a sufficiently large combination of source and receiver pixel positions, there may be sufficient redundancy in the geometry of rays passing through the sample to refine the estimate for the amplitude $\bar{n}_0$ in Eq. 8 based on the sample-present measurements only (see, e.g., [5] and references therein).

Of course, for tomographic purposes, the critical dependence of Eq. 8 is on the variation of M with the precise fluorescent photon path, which could be quite complex due to discontinuities of μ across sample constituent boundaries. As alluded to above, the coordinate system is defined by the fixed function $\mu(\mathbf{x})$, while all other parameters, including the detailed geometry of the function Φ_S , vary with the measurement.

The aim is to reconstruct $\mu(\mathbf{x})$ from an appropriately large set of line integral estimates $\{M(\mathbf{x}_S^0, \mathbf{x}_R^0; E)\}$. These are extracted from $n_{\text{meas}}(\mathbf{x}_S^0, \mathbf{x}_R^0; E)$, after subtraction of the continuous background, and properly accounting for photon Poisson statistics in the low count regime (see Section 3).

2.2 Simplified form

Equation 8 implies, in general, a very complicated exponential relationship between the source and receiver profiles and the key line integrals. However, there are natural conditions under which the averages over these profiles can be moved directly to the line integrals themselves. Specifically, the experimental design and X-ray energies are deliberately selected so that $e^{-M(\mathbf{x}_S, \mathbf{x}_R; E)}$ is large enough that a reasonable fraction of the $n_0(\mathbf{x}_S, E)$ source photons make it through the sample, generating a correspondingly reasonable fluorescence count rate, distinguishable from the background. Thus, $M(\mathbf{x}_S, \mathbf{x}_R; E) = O(1)$, and we further assume that, for most source-receiver pairs $(\mathbf{x}_S^0, \mathbf{x}_R^0)$, the variation of M with $\mathbf{x}_S - \mathbf{x}_S^0$ and $\mathbf{x}_R - \mathbf{x}_R^0$ is weak:

$$M(\mathbf{x}_S, \mathbf{x}_R; E) \equiv M(\mathbf{x}_S^0, \mathbf{x}_R^0; E) + \delta M(\mathbf{x}_S, \mathbf{x}_R; E), \quad \frac{|\delta M|}{M^0} \ll 1. \quad (10)$$

There are clearly cases where this will fail, such as when a ray happens to run for a long distance parallel and immediately adjacent to a straight metal–nonmetal boundary. However, this will likely be a rare occurrence: much more commonly, slight adjustments of a ray will produce similarly small changes in the material profile along the ray, hence only slight adjustments to the net line integral. Note that if this were not the case, then the experimental resolution requirements are likely not being met, and the inversion will not produce an accurate result.

Proceeding under the assumption that the experimental design ensures validity of Eq. 10, to high accuracy one obtains as follows:

$$\begin{aligned} \langle e^{-M(\mathbf{x}_S, \mathbf{x}_R; E)} \rangle_0 &\approx e^{-M(\mathbf{x}_S^0, \mathbf{x}_R^0; E)} \\ &\quad \times [1 - \langle \delta M(\mathbf{x}_S, \mathbf{x}_R; E) \rangle_0] \\ &\approx e^{-M(\mathbf{x}_S^0, \mathbf{x}_R^0; E) - \langle \delta M(\mathbf{x}_S, \mathbf{x}_R; E) \rangle_0} \\ &= e^{-\bar{M}(\mathbf{x}_S^0, \mathbf{x}_R^0; E)} \end{aligned} \quad (11)$$

in which

$$\begin{aligned} \bar{M}(\mathbf{x}_S^0, \mathbf{x}_R^0; E) &= \langle M(\mathbf{x}_S, \mathbf{x}_R; E) \rangle_0 \\ &= \int d\mathbf{x}_S \Phi_S(\mathbf{x}_S - \mathbf{x}_S^0; E) \\ &\quad \times \int_{A_R} \frac{d\mathbf{a}_R \cdot \hat{\mathbf{e}}_{SR}}{\mathbf{A}_R \cdot \hat{\mathbf{e}}_{SR}} \\ &\quad \times \int_0^{|\mathbf{x}_R - \mathbf{x}_S|} \mu[\mathbf{x}(s); E] ds \end{aligned} \quad (12)$$

is the measurement average of the line integral. Substituting (11) into (8), the mean photon count rate follows in the form

$$\begin{aligned} \bar{n}_{\text{meas}}(\mathbf{x}_S^0, \mathbf{x}_R^0; E) &\approx \frac{\Delta \Omega_{SR}}{4\pi} \bar{n}_0(\mathbf{x}_S^0) e^{-\bar{M}(\mathbf{x}_S^0, \mathbf{x}_R^0; E)} \\ &\quad + \bar{n}_B(\mathbf{x}_S^0, \mathbf{x}_R^0; E) \Delta E. \end{aligned} \quad (13)$$

Higher order terms $\langle \delta M^2 \rangle_0$ are neglected. Equation 13 conveniently allows one to treat the measured photon count as a single Poisson process involving a “cone volume” average of rays, rather than a complex superposition of Poisson processes for each individual ray in the cone. This greatly simplifies the inversion scheme. For an actual microchip, there will be structure on many length scales, requiring a multi-scale approach to estimate μ (or, more specifically, the atomic densities n_A) constrained by available prior information.

2.3 Measurement Imperfections

In the above discussion, we have assumed perfect knowledge of the source positions $\mathbf{x}_S$ and the associated function Φ_S . These will be impacted by electron beam wandering and by sample stage translation and rotation uncertainties. We have also not yet accounted for detector uncertainties, such as pixel cross-talk and multiple photon absorption within a detection time window, that will generate, respectively, photon count and energy estimate errors.

For typical CT scans which use scintillator detectors, rather than being directly absorbed the X-ray photon creates a cascade of visible light photons that are then converted to an electronic signal. Scattering of the photons between pixels can be an important blurring effect.

In contrast, the TES detector pixel fully absorbs an X-ray photon, generating an equilibrium heating effect that translates to a superconducting element conductivity change that in turn measures the photon energy [6]. The detailed equilibration process is very complicated, and might also deposit energy in more than one pixel. However, the energy deposited in each pixel, while summing to E , would be less than E , hence would appear as part of the background Bremsstrahlung. This leads to a fluorescent X-ray undercount rather than a spreading of the count [2]. Such processes might then be modeled in the form [1, 2]

$$\bar{n}_{\text{meas}}(\mathbf{x}_S^0, \mathbf{x}_R^0; E) \rightarrow \varepsilon(\mathbf{x}_R^0, E) \bar{n}_{\text{meas}}(\mathbf{x}_S^0, \mathbf{x}_R^0; E) \quad (14)$$

in which $\varepsilon(\mathbf{x}_R^0, E) \leq 1$ is an overall efficiency factor for the detector pixel defined by $\mathbf{x}_R^0$. This is significantly simpler than the scintillator blurring model that replaces the scalar ε by a matrix $\mathbf{B}$ that mixes the count among nearby pixels [2].

3 Tomographic Inversion

The MBIR result for the absorption function $\mu(\mathbf{x})$ is based on minimization of an appropriate objective function

$$\hat{\mu}(\mathbf{n}) = \arg \min_{\mu} \Phi(\mu, \mathbf{n}) \quad (15)$$

in which the array μ now represents the absorption function values over an M element sample grid, and the array $\mathbf{n}$ represents the (integer photon count) data. The function Φ takes a penalized likelihood (PL) form

$$\Phi(\mu, \mathbf{n}) = L(\mu, \mathbf{n}) + \beta R(\mu), \quad (16)$$

with log-likelihood term L , quantifying the photon count Poisson statistics, and regularization term R which encompasses all sample prior knowledge. The latter includes prior constraints on the decomposition (4), such as material step-wise uniformity (with sharp, flat interfaces), metal interconnect geometry, and expected metal types in various sample regions. The relative weight β controls the balance between data fidelity and regularization and its optimal choice must be part of an investigation using various known test samples.

For the present problem, the data consist of independent photon count measurements for each source-receiver pair, with mean count

$$\begin{aligned} \bar{n}_l &= \tau_l n_{\text{meas}}(\mathbf{x}_{S,l}^0, \mathbf{x}_{R,l}^0; E) \\ &\equiv b_l \langle e^{-x_l} \rangle_0 + r_l \end{aligned} \quad (17)$$

where τ_l is the dwell time, and we define the following:

$$\begin{aligned} b_l &= \tau_l \frac{\Delta \Omega_{SR}}{4\pi} \bar{n}_0(\mathbf{x}_{S,l}^0) \\ r_l &= \tau_l \bar{n}_B(\mathbf{x}_{S,l}^0, \mathbf{x}_{R,l}^0; E) \Delta E \\ x_l &= M(\mathbf{x}_{S,l}, \mathbf{x}_{R,l}; E). \end{aligned} \quad (18)$$

These are, respectively, the source mean photon count (determined by target and e-beam properties), the mean background count (obtained, e.g., by a smooth fit over a range of energies surrounding the fluorescent energy), and the attenuation factor. The first two are, for now, assumed accurately known, while the latter contains the μ -dependence targeted by the inversion. For simplicity, additional dependence on beam incident direction $\hat{\mathbf{e}}_S$, beam geometry, etc., has been suppressed from the notation, and extensions to problems where $\bar{n}_0$ is also uncertain [5] will not be treated here.

The Poisson likelihood takes the form of an independent sum over all measurements,

$$L(\mu, \mathbf{n}) = \sum_{l=1}^N h_l(n_l, \bar{n}_l), \quad (19)$$

in which

$$\begin{aligned} h_i &= -\ln \left(\frac{\bar{n}_l^{n_l} e^{-\bar{n}_l}}{n_l!} \right) \\ &= \bar{n}_l - n_l \ln(\bar{n}_l) + \ln(n_l!), \end{aligned} \quad (20)$$

with n_l being the actual measured (integer) photon count. Recall here that the dependence on μ resides, via (1) and (2), in the mean count $\bar{n}_l = \bar{n}_l[\mu]$.

3.1 Simplified Form Likelihood

In general, the μ -dependence of $\bar{n}_l$ is complicated (being a superposition of exponentials of linear functionals of μ), but with the approximation (10), Eq. (13) allows one to simplify (17) to the form

$$\bar{n}_l(x_l) = b_l e^{-\bar{x}_l} + r_l, \quad \bar{x}_l \equiv \langle x_l \rangle_0. \quad (21)$$

The key simplification here is that μ now enters through the *single* linear functional $\bar{x}_l = \bar{x}_l[\mu]$. Explicitly, upon discretizing the sample one may write an individual line integral in the form

$$M(\mathbf{x}_S, \mathbf{x}_R; E) = \sum_{j=1}^M w_j(\mathbf{x}_S, \mathbf{x}_R) \mu_j \quad (22)$$

in which $w_j(\mathbf{x}_S, \mathbf{x}_R)$ is the length of the intersection of the line segment connecting $\mathbf{x}_S$ to $\mathbf{x}_R$ with voxel j . It follows that one may write

$$\bar{x}_l = \sum_j \bar{w}_{lj} \mu_j \equiv \bar{\mathbf{w}}_l \cdot \boldsymbol{\mu} \quad (23)$$

in which the (non-negative) matrix elements

$$\begin{aligned}\bar{w}_{lj} &= \langle w_j(\mathbf{x}_{S,l}, \mathbf{x}_{R,l}) \rangle_0 \\ &\equiv \int d\mathbf{x}_S \Phi_S(\mathbf{x}_S - \mathbf{x}_{S,l}^0; E) \\ &\quad \times \int_{A_{R,l}} \frac{d\mathbf{a}_{R,l} \cdot \hat{\mathbf{e}}_{SR,l}^0}{\mathbf{A}_{R,l} \cdot \hat{\mathbf{e}}_{SR,l}^0} w_j(\mathbf{x}_S, \mathbf{x}_R)\end{aligned}\quad (24)$$

are weighted averages of the intersection lengths over the paths associated with measurement l (characterized by source spot center $\mathbf{x}_{S,l}^0$ and receiver pixel center $\mathbf{x}_{R,l}^0$).

The function $h_l[n_l, \bar{n}_l(x)]$ has a unique minimum at

$$b_l e^{-x} + r_l = n_l \Rightarrow x = \ln\left(\frac{b_l}{n_l - r_l}\right), \quad (25)$$

but is not a convex function of x if $r_l > 0$ —unfortunately ruling out certain numerically efficient objective function minimization algorithms. The result (25) simply states that the most likely absorption model x is such that the measured photon count is precisely the expected count: $n_l = \bar{n}_l(x)$. Except for statistical outliers, the measured photon count will typically lie in the range $b_l + r_l > n_l > r_l$, yielding the physically required positive minimum, $x > 0$.

Note that a simplified measurement model might forgo the average in Eqs. 17 or 24, using instead the single set of line segment lengths

$$w_{lj}^0 \equiv w_j(\mathbf{x}_{S,l}^0, \mathbf{x}_{R,l}^0), \quad (26)$$

associated with the single line connecting the measurement centroids. The result is then an inversion that neglects “blurring” effects in the real data. In principle, for sufficiently large photon counts, one may expect there to be a noticeable difference between images derived from the two different models. As long as parameters $\{b_l, r_l, \mathbf{x}_{S,l}, \mathbf{x}_{R,l}\}$, along with the associated fluorescence weighting functions $\{\Phi_{S,l}\}$ (which depend on the measurement through the local fluorescent layer structure and orientation) are accurately known, the blur-aware model should produce an unblurred image from the blurred data, whereas the mismatched model will produce a blurred image from blurred data.

4 Image Quality Metrics: Smooth Reconstructions

We consider first tomographic reconstructions based on smoothly varying targets, controlled by L^2 -type regularizations. This rules out piecewise constant type prior constraints on which (4) is based. We will generalize the theory to include the latter in subsequent sections.

4.1 Spatial Resolution Metrics

In what follows, we assume the existence of a rapidly converging algorithm that produces $\hat{\mu}[\mathbf{n}]$ for given fixed data array $\mathbf{n}$ and underlying statistical signal forward model defined by Eqs. 17–20.

The simplest image quality metrics are based on “noise free” data. Thus, let $\hat{\mu}(\bar{\mathbf{n}})$ be the solution (15)

$$\hat{\mu}(\bar{\mathbf{n}}) = \arg \min_{\mu} \Phi(\mu, \mathbf{n}) \Big|_{\mathbf{n}=\bar{\mathbf{n}}(\mu)} \quad (27)$$

obtained by replacing the data values by the mean values, $n_i \rightarrow \bar{n}_i[\mu]$, following the minimization (i.e., the replacement is *not* $\mathbf{n} = \bar{\mathbf{n}}(\hat{\mu})$, which would entail a minimization that includes the μ -dependence of $\bar{\mathbf{n}}$). Since n_i appears linearly in the μ -dependent terms in Eq. 20 (the $n_i!$ term is a trivial normalization that plays no role in the minimization (15)), the evaluation of Eq. 27 for continuous $\bar{\mathbf{n}}$ is straightforward.

Equation 27 defines an estimator $\hat{\mu} = \hat{\mu}[\bar{\mathbf{n}}(\mu)]$ as a function of the true μ . An obvious measure of image quality is therefore how closely the former reproduces the latter. A way to quantify this is through the set of point spread functions

$$\begin{aligned}\mathbf{P}_j &= \lim_{\epsilon \rightarrow 0} \frac{\hat{\mu}[\bar{\mathbf{n}}(\mu + \epsilon \hat{\mathbf{e}}_j^{(\mu)})] - \hat{\mu}[\bar{\mathbf{n}}(\mu)]}{\epsilon} \\ &= \frac{\partial \hat{\mu}[\bar{\mathbf{n}}(\mu)]}{\partial \mu_j} = \frac{\partial \bar{\mathbf{n}}(\mu)}{\partial \mu_j} \cdot \nabla_{\mathbf{n}} \hat{\mu}[\bar{\mathbf{n}}(\mu)],\end{aligned}\quad (28)$$

in which $\epsilon[\hat{\mathbf{e}}_j^{(\mu)}]_l = \epsilon \delta_{lj}$ perturbs the absorption on the single voxel j . These measure the change in the estimated $\hat{\mu}$ associated with a change in the actual μ (under the assumption of noise-free data). In a perfect world, $\hat{\mu} = \mu$, and one obtains $[\mathbf{P}_j]_i = \delta_{ij}$ —the inversion result precisely tracks the input value. More realistically, the support of $\mathbf{P}_j$ will spread to nearby voxels, and optimal measurement design is aimed at minimizing this spread.

The last line of Eq. 28 naturally factors the point spread function into the product of (a) a vector quantifying the sensitivity of the data to absorption changes at the single site $\mathbf{x}_j$, with (b) a matrix quantifying the sensitivity of the inversion to changes in the data. Using Eqs. 17 and 18, one obtains for term (a):

$$\frac{\partial \bar{n}_l(\mu)}{\partial \mu_j} = -b_l \langle w_{lj} e^{-x_l} \rangle_0. \quad (29)$$

The result is nonzero only if $\bar{w}_{lj} > 0$, i.e., the ray intersects voxel j within the support of the average (24).

The matrix term (b) quantifies the degree to which the set of measurements (together with the regularization conditions) can be used to resolve a unique signature from voxel j . Resolution will improve as one increases the number of line integrals passing through $\mathbf{x}_j$ from as diverse

a set of directions as possible. An explicit expression is obtained from the minimum condition

$$\nabla_{\mu} \Phi(\hat{\mu} + \delta\mu, \bar{\mathbf{n}} + \delta\mathbf{n}) = 0 \quad (30)$$

in the neighborhood of the given solution $\hat{\mu}[\bar{\mathbf{n}}]$ at $\delta\mu = 0$, $\delta\mathbf{n} = 0$. We emphasize here again that ∇_{μ} operates only on the first argument of Φ , not on the implicit μ -dependence of $\bar{\mathbf{n}}$. To linear order, one obtains

$$(\delta\mu \cdot \nabla_{\mu}) \nabla_{\mu} \Phi_0 + (\delta\mathbf{n} \cdot \nabla_{\mathbf{n}}) \nabla_{\mu} \Phi_0 = 0, \quad (31)$$

in which the subscript 0 indicates that the second derivatives are evaluated at $\delta\mu = 0$, $\delta\mathbf{n} = 0$. The smooth reconstruction assumption being made in this section is equivalent here to the existence of the derivatives in Eq. 31, and as a consequence one obtains

$$\nabla_{\mathbf{n}} \hat{\mu} = -(\nabla_{\mu} \nabla_{\mu} \Phi_0)^{-1} (\nabla_{\mu} \nabla_{\mathbf{n}} \Phi_0). \quad (32)$$

Note here that $\nabla_{\mu} \nabla_{\mu} \Phi_0$ is a square matrix, hence with nominally well-defined inverse, while $\nabla_{\mu} \nabla_{\mathbf{n}} \Phi_0$ is in general not square. From Eqs. 16–20, one obtains

$$\begin{aligned} \nabla_{\mu} \nabla_{\mu} \Phi &= \sum_l \left[\left(1 - \frac{n_l}{\bar{n}_l} \right) \nabla_{\mu} \nabla_{\mu} \bar{n}_l \right. \\ &\quad \left. + \frac{n_l}{\bar{n}_l^2} (\nabla_{\mu} \bar{n}_l) (\nabla_{\mu} \bar{n}_l)^T \right] + \mathbf{R} \\ \nabla_{\mu} \nabla_{\mathbf{n}} \Phi &= - \sum_l \frac{n_l}{\bar{n}_l} (\nabla_{\mu} \bar{n}_l) \mathbf{e}_l^{(n)} \end{aligned} \quad (33)$$

in which $[\mathbf{e}_l^{(n)}]_m = \delta_{lm}$ is the unit vector along data dimension l , while

$$\mathbf{R} = \nabla_{\mu} \nabla_{\mu} R \quad (34)$$

is the corresponding second derivative matrix derived from the regularization term. It is precisely the existence of the matrix $\mathbf{R}$ that defines a smooth reconstruction [1, 2]. A typical choice is a quadratic form for R , giving rise to a constant, positive definite matrix $\mathbf{R}$. The mean photon count second derivatives are given by

$$\frac{\partial^2 \bar{n}_l(\mu)}{\partial \mu_i \partial \mu_j} = b_i \langle w_{il} w_{jl} e^{-x_l} \rangle_0. \quad (35)$$

Finally, setting $\mu \rightarrow \hat{\mu}[\bar{\mathbf{n}}[\mu]]$ and $\mathbf{n} \rightarrow \bar{\mathbf{n}}[\mu]$, one obtains

$$\begin{aligned} \nabla_{\mu} \nabla_{\mu} \Phi_0 &= \sum_l \left[\left(1 - \frac{\bar{n}_l(\mu)}{\bar{n}_l(\hat{\mu})} \right) (\nabla_{\mu} \nabla_{\mu} \bar{n}_l)_0 \right. \\ &\quad \left. + \frac{\bar{n}_l(\mu)}{\bar{n}_l(\hat{\mu})^2} (\nabla_{\mu} \bar{n}_l)_0 (\nabla_{\mu} \bar{n}_l)_0^T \right] + \mathbf{R} \\ \nabla_{\mu} \nabla_{\mathbf{n}} \Phi_0 &= - \sum_l \frac{\bar{n}_l(\mu)}{\bar{n}_l(\hat{\mu})} (\nabla_{\mu} \bar{n}_l)_0 \mathbf{e}_l^{(n)} \end{aligned} \quad (36)$$

in which the subscript 0 in the $\bar{n}_l$ -gradient terms denotes substitution $\mu \rightarrow \hat{\mu}$ inside x_l in Eqs. 29 and 35.

For sufficiently high-quality data, one may expect $\bar{n}_l[\mu]/\bar{n}_l[\hat{\mu}] \simeq 1$, and to leading order (36) simplifies to

$$\begin{aligned} \nabla_{\mu} \nabla_{\mu} \Phi_0 &\approx \sum_l \frac{1}{\bar{n}_l(\mu)} (\nabla_{\mu} \bar{n}_l) (\nabla_{\mu} \bar{n}_l)^T + \mathbf{R} \\ \nabla_{\mu} \nabla_{\mathbf{n}} \Phi_0 &\approx -\nabla_{\mu} \bar{\mathbf{n}}, \end{aligned} \quad (37)$$

which involve only the first derivative (29), and all quantities are evaluated at the input value μ . In this way, under conditions where the image is expected to be reasonably accurate, correspondingly accurate image quality metrics may be derived from the forward model alone, without actually needing to first derive the tomographic inversion $\hat{\mu}$.

4.2 Local Noise Metrics

The next step is to explore the effects of noise, measuring image quality in the presence of both model imperfection and measurement noise. Adopting the linear approximation,

$$\hat{\mu}(\mathbf{n}) = \hat{\mu}(\bar{\mathbf{n}}) + [\mathbf{n} - \bar{\mathbf{n}}(\mu)] \cdot \nabla_{\mathbf{n}} \hat{\mu}[\bar{\mathbf{n}}(\mu)], \quad (38)$$

the effects of noise may be quantified by the covariance

$$\mathbf{C}(\hat{\mu}) = \nabla_{\mathbf{n}} \hat{\mu}[\bar{\mathbf{n}}(\mu)]^T \mathbf{C}(\bar{\mathbf{n}}) \nabla_{\mathbf{n}} \hat{\mu}[\bar{\mathbf{n}}(\mu)] \quad (39)$$

in which

$$\mathbf{C}(\bar{\mathbf{n}}) = \langle [\mathbf{n} - \bar{\mathbf{n}}(\mu)] [\mathbf{n} - \bar{\mathbf{n}}(\mu)]^T \rangle \quad (40)$$

is the underlying photon count covariance. In our model, the Poisson statistics of each measurement are independent and $\mathbf{C}[\bar{\mathbf{n}}]$ is therefore diagonal, with each measurement variance equal to the mean:

$$\mathbf{C}[\bar{\mathbf{n}}]_{lm} = \langle (n_l - \bar{n}_m)(n_m - \bar{n}_m) \rangle = \bar{n}_l \delta_{lm} \approx n_l \delta_{lm}. \quad (41)$$

The local image covariance $\mathbf{C}(\hat{\mu})$ is in general non-diagonal since the matrix $\nabla_{\mathbf{n}} \hat{\mu}[\bar{\mathbf{n}}(\mu)]$, the same as that appearing in Eqs. 28 and 32, is nonlocal. It follows that sensitivity of a particular photon count to μ spreads to the neighborhood of the corresponding source-receiver ray.

4.3 Simplified Form Image Quality Metrics

For the simplified form Eq. 21, with Eq. 23, one obtains (via (25))

$$\bar{n}_l(\mu) = b_l e^{-\bar{\mathbf{w}}_l \cdot \mu} + r_l, \quad (42)$$

which then produces

$$\begin{aligned} \frac{\partial \bar{n}_l(\mu)}{\partial \mu_j} &= -\bar{w}_{lj} b_l e^{-\bar{\mathbf{w}}_l \cdot \mu} = -\bar{w}_{lj} (\bar{n}_l - r_l) \\ &\approx -\bar{w}_{lj} (n_l - r_l) \\ \frac{\partial^2 \bar{n}_l(\mu)}{\partial \mu_i \partial \mu_j} &= \bar{w}_{li} \bar{w}_{lj} (\bar{n}_l - r_l) \end{aligned} \quad (43)$$

and depends on a combination of the (average) line integral sensitivity $\bar{\mathbf{w}}_l$ of measurement l to voxels i, j and the overall expected non-background photon count $(\bar{n}_l - r_l)$.

Once again, with high-quality data, one may use $\bar{n}_l[\hat{\mu}]$ interchangeably with $\bar{n}_l[\mu]$, avoiding the need to first reconstruct $\hat{\mu}$. Moreover, if the photon counts are sufficiently high, one will have $n_l/\bar{n}_l \approx 1$, and the quality metrics may then be computed directly from the data with the replacement $\bar{\mathbf{n}} \rightarrow \mathbf{n}$ [2]: the simplified forms require only the blur-averaged geometrical parameters $\bar{\mathbf{w}}_l$, measured photon counts n_l , and data-derived background counts r_l without any direct reference to μ or $\hat{\mu}$.

Correspondingly, from a simulation point of view, one may investigate the quality of a proposed experimental setup using the forward model alone to generate the values $\bar{\mathbf{n}}(\mu)$ (along with background count estimates) from a given target model μ . One may either use these values for $\mathbf{n}$, or add further realism by generating simulated Poisson-distributed counts from these computed average counts. One is not required to go through the more complicated inversion step of first deriving $\hat{\mu}[\bar{\mathbf{n}}(\mu)]$ from μ .

4.4 Numerical Considerations

The matrix inversion in Eq. 32 is perhaps the most numerically intensive step in the computation of the point spread functions (28). Under certain conditions, this step may be sped up. An important case is that of approximate translation invariance, where Fourier transforms can be used to speed up the inversion tremendously. Thus, let

$$A_{ij} = A(\mathbf{x}_i, \mathbf{x}_j) \quad (44)$$

be a matrix that is a function in both indices of a coordinate on a rectangular grid. One may always rewrite this as a function of the difference and center coordinates,

$$A_{ij} = \tilde{A}\left(\mathbf{x}_i - \mathbf{x}_j; \frac{\mathbf{x}_i + \mathbf{x}_j}{2}\right). \quad (45)$$

However, we suppose now that $\tilde{A}$ is local in the first argument, effectively vanishing outside of some range r_0 , $A_{ij} \rightarrow 0$ for $|\mathbf{x}_i - \mathbf{x}_j| > r_0$, while essentially constant over this same range in the second argument.

The objective is to solve a matrix equation of the form

$$\sum_j A_{ij} B_{jl} = C_{il} \Rightarrow B_{il} = \sum_j \left[A^{-1} \right]_{ij} C_{jl}, \quad (46)$$

in which we define

$$B_{il} = B(\mathbf{x}_i, n_l), \quad C_{il} = C(\mathbf{x}_i, n_l), \quad (47)$$

with indices lying in the distinct spatial grid and photon measurement spaces. For the present problem, we identify

$$\begin{aligned} A(\mathbf{x}_i, \mathbf{x}_j) &\equiv \frac{\partial^2 \Phi_0}{\partial \mu(\mathbf{x}_i) \partial \mu(\mathbf{x}_j)} \\ B(\mathbf{x}_i, n_l) &= \frac{\partial \hat{\mu}(\mathbf{x}_i)}{\partial n_l} \\ C(\mathbf{x}_i, n_l) &= \frac{\partial \Phi_0}{\partial \mu(\mathbf{x}_i) \partial n_l}. \end{aligned} \quad (48)$$

Under the slowly varying condition, one may then derive an approximate solution to Eq. 46 using Fourier transform “windowing.” Thus, about each given point $\mathbf{x}$, consider a box of diameter $2r_0$, and define the local Fourier transforms

$$\begin{aligned} \hat{A}(\mathbf{q}, \mathbf{r}) &= \frac{1}{M} \sum_i' e^{-i\mathbf{q} \cdot \mathbf{x}_i} \tilde{A}(\mathbf{x}_i, \mathbf{x}), \\ \hat{B}_l(\mathbf{q}; \mathbf{x}) &= \frac{1}{M} \sum_i' e^{-i\mathbf{q} \cdot \mathbf{x}_i} B(\mathbf{x} + \mathbf{x}_i, n_l) \\ \hat{C}_l(\mathbf{q}; \mathbf{x}) &= \frac{1}{M} \sum_i' e^{-i\mathbf{q} \cdot \mathbf{x}_i} C(\mathbf{x} + \mathbf{x}_i, n_l) \end{aligned} \quad (49)$$

in which M is the number of points in the box, and the prime on the sums indicates restriction to the corresponding box centered at the origin. The wave vector $\mathbf{q}$ ranges over the box domain reciprocal lattice, and describes the more rapid variation of A and B on scales smaller than r_0 . Note that since $\tilde{A}$ is short range in its first argument, the sum restriction is essentially redundant.

Applying the local Fourier transform to both sides of (46), one obtains

$$\begin{aligned} \hat{C}_l(\mathbf{q}; \mathbf{x}) &= \sum_j' e^{-i\mathbf{q} \cdot \mathbf{x}_j} B(\mathbf{x} + \mathbf{x}_j, n_l) \sum_i' e^{-i\mathbf{q} \cdot (\mathbf{x}_i - \mathbf{x}_j)} \\ &\quad \times \tilde{A}\left(\mathbf{x}_i - \mathbf{x}_j; \mathbf{x} + \frac{\mathbf{x}_i + \mathbf{x}_j}{2}\right) \\ &\approx \hat{A}(\mathbf{q}; \mathbf{x}) \sum_j' e^{-i\mathbf{q} \cdot \mathbf{x}_j} B(\mathbf{x}_j, n_l) \\ &= \hat{A}(\mathbf{q}; \mathbf{x}) \hat{B}_l(\mathbf{q}; \mathbf{x}) \end{aligned} \quad (50)$$

In the approximate equality, we have neglected the $\mathbf{x}_i - \mathbf{x}_j$ dependence of the second argument of $\tilde{A}$ and thereby succeeded in proving a local version of the convolution theorem. The matrix inversion now reduces to an algebraic division of both sides by $\hat{A}$ and a subsequent inverse Fourier transform, and one obtains the desired result

$$B(\mathbf{x} + \mathbf{x}_i, n_l) \approx \sum_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{x}_i} \frac{\hat{C}_l(\mathbf{q}; \mathbf{x})}{\hat{A}(\mathbf{q}; \mathbf{x})}. \quad (51)$$

Here, the result is restricted to the local box neighborhood of a chosen point $\mathbf{x}$. The global solution is obtained by varying $\mathbf{x}$ over the appropriate global set of box centers.

4.4.1 Consistency Conditions

It remains to discuss the conditions under which one indeed expects $\tilde{A}$ to obey the desired conditions.

First, the regularization term R is typically local, e.g., depending on near-neighbor differences $\mu(\mathbf{x}_i) - \mu(\mathbf{x}_j)$. Often a quadratic regularization is used [2], in which case $\mathbf{R}$ is a constant, near-diagonal matrix (or perhaps a slowly varying set of near-diagonal quadratic coefficients, depending on prior knowledge of the target), and the desired conditions indeed hold because such a regularization biases the solution toward relatively smooth $\hat{\mu}(\mathbf{x})$.

Similarly, if the bias is toward smooth, slowly varying $\hat{\mu}(\mathbf{x})$, the likelihood term (19) will be a smooth function of the mean counts $\bar{\mathbf{n}}(\hat{\mu})$. However, each independent count $\bar{n}_l$ depends on a narrow cylinder of voxels connecting source and receiver and is hence strongly nonlocal. In particular, the second derivative (35) will be nonzero for any pair $\mathbf{x}_i, \mathbf{x}_j$ lying within the cylinder defined by $\bar{n}_l$.

On the other hand, the sum over l in Eq. 19 contains many cylinders, and one expects high-resolution images to emerge only when each voxel is intersected by many cylinders. In this case for widely separated $\mathbf{x}_i, \mathbf{x}_j$, one expects a small number of cylinders to pass through both, and a correspondingly small number of terms will contribute to the l -sum in the first equalities in Eqs. 36 or 37. In contrast, for $\mathbf{x}_i, \mathbf{x}_j$ within a cylinder diameter, many terms will contribute. It follows that $\partial^2 L / \partial \mu_i \partial \mu_j$ will be strongly peaked about the diagonal $\mathbf{x}_i = \mathbf{x}_j$, and the locality property emerges.

5 Generalized Image Quality Metrics

We next consider more general classes of regularization terms R . The structure of L remains the same as before, and we continue to assume that many l -cylinders pass through each voxel. Thus, $\partial^2 L / \partial \mu_i \partial \mu_j$ continues to be near diagonal. More problematical is the structure (4) in which the densities $n_A(\mathbf{x})$ are piecewise slowly varying, but with sharp interfaces between, further biased, e.g., toward metal interconnect rectilinear geometry. The regularization term will be relatively agnostic to the position of the interface, but sharp interfaces dictate an L^1 rather than L^2 regularization. The second derivative of R will therefore be very singular when evaluated at $\hat{\mu}(\mathbf{x})$. Moreover, strong rectilinear constraints may lead to global shifts in an interconnect position with small change in $\mathbf{n}$. For example, for cylinders aligned along a preferred interconnect direction, the value

of n_l will have a large jump as the cylinder crosses from one side of a metal interface to the other.

5.1 Illustrative Regularization

Let the target consist of K different materials/compounds with absorption $\{\mu(\kappa)\}_{\kappa=1}^K$. Each target voxel is then assigned a material index $M_i \in \{1, 2, \dots, K\}$ with absorption $\mu_i = \mu(M_i)$. A simple regularization term might take the Potts model form

$$R(\mathbf{M}) = - \sum_{\kappa=1}^K h_{\kappa} \sum_i \delta_{M_i, \kappa} - \sum_{\kappa=1}^K J_{\kappa} \sum_{\langle i, j \rangle} \delta_{M_i, \kappa} \delta_{M_j, \kappa}, \quad (52)$$

in which the fields h_{κ} are used to control the fractional area of each material type, and the coupling constants $J_{\kappa} > 0$ favor nearest neighbor voxels $\langle i, j \rangle$ being of the same material type. Both could also be made slow functions of position in order to encode prior information on material types in different regions of the target.

Additional terms may be included that favor particular material boundary orientations (e.g., rectilinear). For example, if a, b, c, d are counter-clockwise (starting from the first quadrant) labels of a 2×2 plaquette centered on some point i , with corresponding material labels $M_a(i), M_b(i), M_c(i), M_d(i)$, then the terms in

$$\begin{aligned} R_4(\mathbf{M}) = & - \sum_{\kappa} L_{\kappa} \sum_i [\delta_{M_a(i), \kappa} \delta_{M_b(i), \kappa} \\ & \times (1 - \delta_{M_c(i), \kappa})(1 - \delta_{M_d(i), \kappa}) \\ & + \delta_{M_c(i), \kappa} \delta_{M_d(i), \kappa} \\ & \times (1 - \delta_{M_a(i), \kappa})(1 - \delta_{M_b(i), \kappa})] \end{aligned} \quad (53)$$

are nonzero only if sites a, b are both the same material but different from c, d , or vice versa, hence favoring horizontal interfaces. Additional terms, for example, could be introduced favoring certain interface types (e.g., metal–insulator).

5.2 Properly Designed Image Quality Metrics

By balancing the data-driven likelihood term L against this alternative type of regularization term, characteristic material solutions $\hat{\mathbf{M}}$ will emerge with certain types of enforced geometries (e.g., piecewise constant, preferred shape and orientation)—notional geometries are illustrated in Fig. 2. The size of the coefficients $h_{\kappa}, J_{\kappa}, L_{\kappa}, \dots$ are chosen to emphasize different target features in proportion to one's degree of confidence in such prior information. Of course, the more high-quality data one has (the larger L is relative to R), the less impact these coefficients will have. They are most useful under data-starved conditions in which R is able to resolve ambiguities in L in favor of the desired geometry. For example, some violations of geometry rules may in fact be real (due to target

Fig. 2 Illustration of sample geometry benefiting from Potts model-type regularization (52) and (53) of the objective function (16), including resolution varying with position. Different material constituents (e.g., metallic interconnects colored blue and green; insulating materials colored yellow and orange), are generally contiguous and compact in shape. However, the Poisson likelihood term (19) and (20) ensures that the photon count data strongly influence the shape of the boundary to be consistent with the line integral values (2) along the fans of X-ray paths (receiver array not shown). Proper balance between the two sets of terms is influenced by the density and diversity of line integrals and by a physically reasonable choice for the regularization amplitude β

imperfections, unrecognized manufacturing specifications, etc.), and sufficiently high-quality data must be permitted to resolve the discrepancy in favor of L rather than R . One must therefore ensure that the coefficients in R are chosen in such a way that the regularization does not completely dominate the data. Thus, the difference in the likelihood terms, $|L[\mu(\hat{\mathbf{M}}_0), \bar{\mathbf{n}}_0] - L[\mu(\mathbf{M}_0), \bar{\mathbf{n}}_0]|$ should be small: the adjustments in $\mathbf{M}_0$ that significantly reduce R should remain consistent with the data.

The results here are therefore quite different from the continuously varying solutions favored by the L^2 -type regularizations discussed in Section 4. In particular, it no longer makes sense to define the objective function Φ as a continuous function of μ , and to subsequently formulate the sensitivity of the minimum $\hat{\mu}$ in terms of gradients—see Eqs. 28 and 30–32. The regularization smoothing process is unlikely to be sensitive to changes limited to a single site.

Instead, one needs to recognize that small changes in the photon count data could lead to strongly nonlocal changes in $\hat{\mathbf{M}}$, e.g., small lateral shift of an entire metal interconnect (which would be best sensed by a photon count measurement whose source-receiver cylinder of support strongly overlaps the region of this shift). Thus, instead of considering the effects of single local changes in μ , as in Eq. 28, one should probe the effects of more general changes consistent with R . Specifically, let $\mathbf{M}_0$ be a representative physically consistent material geometry (in

the sense that $R(\mathbf{M}_0)$ is small), with corresponding mean photon counts $\bar{\mathbf{n}}_0 = \bar{\mathbf{n}}[\mu(\mathbf{M}_0)]$, and let

$$\hat{\mathbf{M}}_0 = \arg \min_{\mathbf{M}} \Phi(\mathbf{M}, \bar{\mathbf{n}}_0) \quad (54)$$

be the minimizer. An obvious image quality requirement is that $|\hat{\mathbf{M}}_0 - \mathbf{M}_0|$ should be small (in an appropriate norm that, e.g., measures the number of voxels on which the models differ).

Image quality should also be judged on the ability of the photon count data (sparse as it might be) to quantitatively distinguish changes consistent with R . If the data is so starved that such changes have too small an effect on L , then the tomography experiment cannot be expected to yield robust results. To quantify this, consider the class of changes $\delta\mathbf{M}_0$ such that $|R(\hat{\mathbf{M}}_0 + \delta\mathbf{M}_0) - R(\hat{\mathbf{M}}_0)|$ is small. We then demand that L be able to sense such changes—the measurement design must be capable of disambiguating nearby target geometries that are both consistent with R . Defining the reconstruction change $\delta\hat{\mathbf{M}}_0$ by

$$\hat{\mathbf{M}}_0 + \delta\hat{\mathbf{M}}_0 = \arg \min_{\mathbf{M}} \Phi[\mathbf{M}, \bar{\mathbf{n}}(\mathbf{M}_0 + \delta\mathbf{M}_0)], \quad (55)$$

then a more general image quality metric, generalizing (28), is that

$$\begin{aligned} \mathbf{P}(\mathbf{M}_0, \delta\mathbf{M}_0) &= \hat{\mathbf{M}}[\bar{\mathbf{n}}(\mathbf{M}_0 + \delta\mathbf{M}_0)] - \hat{\mathbf{M}}[\bar{\mathbf{n}}(\mathbf{M}_0)] \\ &\equiv \delta\hat{\mathbf{M}}_0 \end{aligned} \quad (56)$$

should be “close” to $\delta\mathbf{M}_0$ (also in the sense of differing on a small number of target pixels). In the present case, $\delta\mathbf{M}_0$ might correspond to a small, smooth translation of a material boundary, and the desire would be that $\delta\hat{\mathbf{M}}_0$ substantially matches this change.

5.2.1 Forward-Model-Only Approximate Formulation

As discussed below (37), it would be desirable to derive reasonably accurate image quality metrics from the forward model alone. Thus, at the end of Section 4.1, we succeeded in estimating the key second derivative matrices appearing in Eq. 32 in terms of the forward photon data $\bar{\mathbf{n}}(\mu)$ alone without reference to the optimal model $\hat{\mu}$. Here, the presence of $R(\mathbf{M})$, which is deliberately sensitive to the discrete nature of $\mathbf{M}$ in order to help select models with desirable piecewise constant geometries, is a complication because the assumption of continuous second derivatives no longer make sense.

An alternative approach taken here is to consider only models $\mathbf{M}_0$ consistent with the geometrical constraints, hence with relatively small values of $R[\mathbf{M}_0]$, and to consider only perturbations $\delta\mathbf{M}_0$ such that $R(\mathbf{M}_0 + \delta\mathbf{M}_0) - R(\mathbf{M}_0)$ is very small, i.e., such that the change in the tomographic estimate $\delta\hat{\mathbf{M}}_0$ is driven by the change $\bar{\mathbf{n}}[\mu(\mathbf{M}_0 + \delta\mathbf{M}_0)] - \bar{\mathbf{n}}[\mu(\mathbf{M}_0)]$ in the photon count data. Thus, we suppose that the measurement protocol is such that the change in Φ is

dominated by the change in L , which remains continuous in μ .

It is convenient to write the mean photon count likelihood term in the form

$$L(\hat{\mu}, \mathbf{n})|_{\mathbf{n}=\bar{\mathbf{n}}(\mu)} \equiv \tilde{L}(\hat{\mu}, \mu), \quad (57)$$

in which one calls out the separate dependencies on the reconstructed and true absorption profiles $\hat{\mu}, \mu$ (both appearing, via (19), through their respective mean photon counts $\bar{\mathbf{n}}(\hat{\mu}), \bar{\mathbf{n}}(\mu)$). One notes as well that a model perturbation $\mathbf{M}_0 \rightarrow \mathbf{M}_0 + \delta\mathbf{M}$ generates an absorption field change

$$\begin{aligned} \delta\mu(\mathbf{M}_0, \delta\mathbf{M}) &\equiv \mu(\mathbf{M}_0 + \delta\mathbf{M}) - \mu(\mathbf{M}_0) \\ &= \sum_i [\mu(M_{0,i} + \delta M_i) - \mu(M_{0,i})] \mathbf{e}_i \end{aligned} \quad (58)$$

which reassigns the absorption values on the support of $\delta\mathbf{M}$. With this notation, the minimum condition for the reconstruction $\hat{\mu}_0(\mu_0)$ now takes the form

$$\tilde{L}(\hat{\mu}_0 + \delta\hat{\mu}, \mu_0) - \tilde{L}(\hat{\mu}_0, \mu_0) = 0 \quad (59)$$

over the subspace of permissible $\delta\hat{\mu} \equiv \delta\hat{\mu}(\hat{\mathbf{M}}_0, \delta\hat{\mathbf{M}})$. Given that the mean photon counts $\bar{\mathbf{n}}(\mu)$ depend continuously on μ , and that one expects only small changes if the support of $\delta\mu$ is small (except for singular cases, which we neglect, where the support of a long narrow change happens to line up exactly with a ray trajectory), one may approximate (59) by

$$(\delta\hat{\mu} \cdot \nabla_1) \tilde{L}(\hat{\mu}_0, \mu_0) = 0, \quad (60)$$

in which $\nabla_{1,2}$ will refer to μ -derivatives with respect to the first and second arguments.

Considering next small changes $\mu_0 \rightarrow \mu_0 + \delta\mu_0$ to the true profile, the generalization of the minimum condition (31) is

$$(\delta\hat{\mu} \cdot \nabla_1)(\delta\hat{\mu}_0 \cdot \nabla_1 + \delta\mu_0 \cdot \nabla_2) \tilde{L}(\hat{\mu}_0, \mu_0) = 0, \quad (61)$$

which is enforced for all permitted independent choices of $\delta\hat{\mu}, \delta\mu_0$, and is to be solved for the induced change $\delta\hat{\mu}_0$ in the reconstruction for each given $\delta\mu_0$. The second derivatives of $\tilde{L}$ are computed as in Section 4.1. As observed below (37), they may be expressed entirely in terms of the forward model parameters alone—one need not perform the tomographic inversion in order to evaluate the image quality.

A reasonable starting point for investigating solutions to Eq. 61 is to evaluate the second derivatives of $\tilde{L}$ at $\hat{\mu}_0 = \mu_0$ (as in Eq. 37), which then also allows one to draw the perturbations $\delta\hat{\mu}, \delta\hat{\mu}_0, \delta\mu_0$ from the same set. Thus, if $\{\delta\mu_a\}_{a=1}^L$ are the permitted common set of small R perturbations, and one defines the $L \times L$ matrices

$$\begin{aligned} L_{ab}^{(1)} &= (\delta\mu_a \cdot \nabla_1)(\delta\mu_b \cdot \nabla_1) \tilde{L}(\mu_0, \mu_0) \\ L_{ab}^{(2)} &= (\delta\mu_a \cdot \nabla_1)(\delta\mu_b \cdot \nabla_2) \tilde{L}(\mu_0, \mu_0), \end{aligned} \quad (62)$$

then Eq. 61 takes the form

$$L_{ac}^{(1)} + L_{ab}^{(2)} \simeq 0 \quad (63)$$

for all a , and for some value c for each choice of b . The use of “ $\simeq$ ” here indicates that perfect equality is unlikely to be achieved when only discrete values of the absorption on a discrete lattice are permitted (compared to the continuous values permitted in the corresponding matrix (31)). Instead, one will likely only obtain approximate correspondence, e.g., of displacements of various interfaces between Potts domains of uniform material. For a perfect measurement, one expects $c = b$, but in general, there will be an imperfect correspondence between the true and reconstructed perturbations.

Acknowledgments This material is based upon work supported by The United States Air Force and Air Force Research Laboratory under Contract FA8650-17-C-9114.

Compliance with Ethical Standards

Disclaimer Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of The United States Air Force and Air Force Research Laboratory.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Schmitt SM, Goodsitt MM, Fessler JA (2017) Fast variance prediction for iteratively reconstructed CT images with locally quadratic regularization. *IEEE Trans Med Imaging* 36(1):17–26
- Wang W, Gang GJ, Siewerdsen JH, Stayman JW (2019) Predicting image properties in penalized-likelihood reconstructions of flat-panel CBCT. *Med Phys* 46(1):65–80
- Irwin KD, Hilton GC (2005) Transition-edge sensors Cryogenic Particle Detection. Springer, pp 81–97
- Salvat F (2015) Penelope-2014: A code system for Monte Carlo simulation of electron and photon transport. In: OECD Workshop Proceedings, Barcelona
- Rezaei A, Defrise M, Nuyts J (2014) ML-Reconstruction for TOF-PET with simultaneous estimation of the attenuation factors. *IEEE Trans Med Imaging* 33:1563–1572
- Ullom JN, Bennett DA (2015) Review of superconducting transition-edge sensors for x-ray and gamma-ray spectroscopy. *Supercond Sci Technol* 28(8):84003–84038

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