

Online Beam Current Estimation in Particle Beam Microscopy

Sheila W. Seidel, *Student Member, IEEE*, Luisa Watkins, *Student Member, IEEE*,
Minxu Peng, *Student Member, IEEE*, Akshay Agarwal, *Member, IEEE*, Christopher Yu, *Member, IEEE*,
and Vivek K Goyal, *Fellow, IEEE*

Abstract—In conventional particle beam microscopy, knowledge of the beam current is essential for accurate micrograph formation and sample milling. This generally necessitates offline calibration of the instrument. In this work, we establish that beam current can be estimated online, from the same secondary electron count data that is used to form micrographs. Our methods depend on the recently introduced time-resolved measurement concept, which combines multiple short measurements at a single pixel and has previously been shown to partially mitigate the effect of beam current variation on micrograph accuracy. We analyze the problem of jointly estimating beam current and secondary electron yield using the Cramér–Rao bound. Joint estimators operating at a single pixel and estimators that exploit models for inter-pixel correlation and Markov beam current variation are proposed and tested on synthetic microscopy data. Our estimates of secondary electron yield that incorporate explicit beam current estimation beat state-of-the-art methods, resulting in micrograph accuracy nearly indistinguishable from what is obtained with perfect beam current knowledge. Our novel beam current estimation could help improve milling outcomes, prevent sample damage, and enable online instrument diagnostics.

Index Terms—electron microscopy, estimation theory, Fisher information, gallium ion beam, helium ion beam, neon ion beam, Neyman Type A distribution, Poisson processes, Touchard polynomials.

I. INTRODUCTION

PARTICLE beam microscopes image samples by detecting the secondary electrons (SEs) expelled from the sample by an incident beam of charged particles. Scanning electron microscopes (SEM) employ an electron beam while newer focused ion beam (FIB) microscopes use a beam composed of heavier ions, such as helium, neon, or gallium. FIB microscopes tend to have higher SE yield, larger depth of field, and finer resolution [1]–[3]. Due to the heavier incident particles, these instruments are also often used to perform milling. Image quality and milling accuracy depend heavily on the ability

This work was supported in part by a Draper Fellowship, a Boston University Clare Boothe Luce Scholar Award, and by the US National Science Foundation under Grant No. 1815896.

S. W. Seidel is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA and is a Draper Scholar with Charles Stark Draper Laboratory, Cambridge, MA 02139 USA (sseidel@bu.edu).

L. Watkins, M. Peng, and V. K. Goyal are with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (luisaw@bu.edu; mxpeng@bu.edu; v.goyal@ieee.org).

A. Agarwal was with the Department of Electrical and Computer Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (akshayag@mit.edu).

C. Yu is with Charles Stark Draper Laboratory, Cambridge, MA 02139 USA (cyu@draper.com).

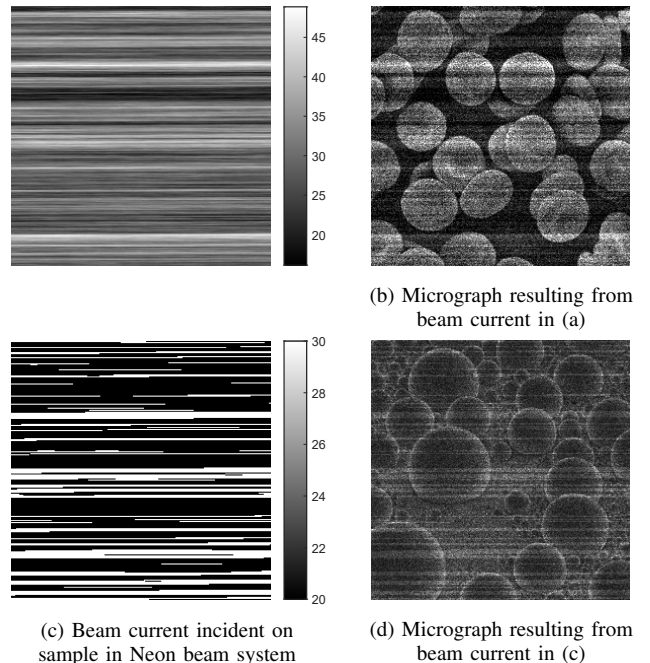


Fig. 1: Synthetic examples of stripe artifacts in simple models for helium and neon beam microscopes. A helium beam current is continuous-valued and slowly varying; a neon beam current has discrete jumps between known values. In (a) and (c), the displayed dose parameter λ is the mean number of incident ions per pixel during one pixel dwell time which is a scaling of the beam current using the ion charge and dwell time.

to maintain a stable beam current. However, contamination within the instrument [4], or age of the source tip, may cause the beam intensity to fluctuate away from the desired setting. When the beam is raster scanned across a sample, these fluctuations give rise to horizontal stripe artifacts in the micrograph. Figure 1a shows the beam current incident on a horizontally raster-scanned sample when the beam has slow, smooth intensity variation as often seen in SEM and helium ion microscope (HIM) systems [5]. Figure 1b depicts the corresponding stripe artifacts that arise when a micrograph is formed under the incorrect assumption of constant beam current. Neon beam microscopes have been less widely adopted because of difficulties in maintaining a stable beam current. Here, the beam current can be modeled as toggling between known values [6], [7], as shown in Figure 1c. The resulting artifacts are shown in Figure 1d. Real-time knowledge of the

beam current could prevent these artifacts, provide the operator an indicator of instrument fitness, and enable the instrument to adjust dwell time to improve milling outcomes and avoid sample damage.

Since existing particle beam microscopes do not measure or estimate beam current, algorithms have been developed to remove stripe content from micrographs post facto. For example, in [5], [8], [9], low-frequency image content is removed, with [9] also including total variation denoising. Several algorithms [10]–[12] have also been developed to reduce the ‘curtaining’ effect that arises due to variations in an ion beam’s milling rate. Unlike the stripe artifacts shown in Figures 1b and 1d, which arise when the image formation algorithm incorrectly assumes a constant beam current, curtaining stripes accurately reflect grooves made in the underlying sample during milling.

In this paper, we establish that beam current can be accurately estimated *online*, from the same SE count data used to form a micrograph, without the use of a calibrated sample. This may seem to be like estimating two quantities from a single noisy measurement. While that would be impossible without additional assumptions, here we take advantage of the time-resolved (TR) measurement concept introduced in [13]. TR measurement divides the dwell time into n shorter sub-acquisitions and can be implemented without changes to the instrument. It enables estimation of the number of incident particles despite the variability in the numbers of SEs generated by the incident particles. This was previously shown to improve micrograph accuracy [13], [14] and to provide a natural robustness to imperfectly known beam current [15], [16]. Another use of multiple short measurements is to reduce blurring due to sample drift [17]. In this work, we go beyond robustness to explicitly estimate the beam current. In addition to improving the accuracy of the estimated SE yield (i.e., producing a better micrograph), this is of interest for control of sample damage, milling accuracy, and instrument diagnostics.

A. Main Contributions

- *An analysis of the joint estimation problem.* The Cramér–Rao bound (CRB) for joint estimation of SE yield and beam current from TR measurement is derived and used to show that under certain conditions, joint estimation is vanishingly more challenging than estimation of SE yield alone.
- *Demonstration of joint estimation of SE yield and beam current at a single pixel.* Proposed estimators are evaluated on synthetic data and compared to the CRB.
- *Joint estimation algorithms that exploit models for beam current variation and inter-pixel correlation.* Causal and non-causal joint estimators are proposed to exploit continuous and discrete Markov models for beam current, and for the continuous case total variation regularization of SE yield is also incorporated. Tested on synthetic microscopy data, the SE yield estimates are shown to beat state of the art methods, and beam current estimates are shown to closely match the ground truth.
- *Numerical methods for Neyman Type A distributions.* Expressions and approximations for derivatives of the

Neyman Type A negative log likelihood in terms of Touchard polynomials are given.

B. Outline

In Section II, we summarize microscope abstractions, measurement models, and basic analyses from [14, Sect. II]. In Section III, we use the Cramér–Rao bound to explore the feasibility of joint estimation. Then, joint estimation at a single pixel is demonstrated in Section IV. At the time scale of pixel-to-pixel scanning, beam current does not vary arbitrarily. Thus, we develop methods to exploit simple Markov models for the beam current. In Section V, motivated by electron and helium ion beams, we consider joint estimation for continuous-valued smoothly varying beam current. In Section VI, motivated by neon ion beams [6], [7], we consider joint estimation where beam current is known to flip back and forth between two known values. Section VII provides concluding comments on the promise of joint estimation in particle beam microscopy.

II. MEASUREMENT MODELS AND BASELINE ESTIMATORS

In this section, we describe our measurement model for the particle beam microscope, assuming direct SE detection; see [14, Sect. II] for additional details. Though not yet common in commercial hardware, direct SE detection provides higher signal-to-noise ratio than employing scintillators and photomultiplier tubes [18]–[21]. The advantages of time-resolved sensing do not depend on direct SE detection, as demonstrated experimentally in [13]. Although our model describes both electron- and ion-beam microscopy, we will refer to incident particles as ions.

A. Conventional Measurement

The incident ion beam may be accurately modeled as a Poisson process [22]. So, at a single pixel with dwell time t , the number of incident ions M is a Poisson random variable with mean $\lambda = \Lambda t$, where Λ is rate of incident ions per unit time. Dose is conventionally defined as the number of incident ions per unit area. Since the absolute spatial and temporal scales are not important in our abstractions or methods, we will refer to λ colloquially as the *dose* and—especially to emphasize when it is unknown and potentially varying—as the *beam current*.¹ The number of detected SEs expelled by the i th incident ion is also modeled as a Poisson random variable: $X_i \sim \text{Poisson}(\eta)$. We refer to η as the sample *SE yield* at that pixel. As discussed in [23], detector quantum efficiency can be nearly 1. To relate the *emitted* SE rates to the *detected* SE rates, one may scale by the inverse of the detector efficiency.

The goal in forming a micrograph is to show how η varies from pixel to pixel. The number of incident ions M is not directly observed; it is at best inferred from the detected SEs.

At each pixel, a conventional microscope measures the sum of all SEs over the dwell time: $Y = \sum_{i=1}^M X_i$. Modeled in

¹The true beam current is *proportional* to λ : $i_B = q\lambda = q\Lambda/t$, where q is the elementary electron charge.

this way, Y is a *Neyman Type A* compound Poisson random variable with probability mass function (PMF) given by

$$P_Y(y; \eta, \lambda) = \frac{e^{-\lambda} \eta^y}{y!} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \quad y = 0, 1, \dots, \quad (1)$$

with mean

$$E[Y] = \lambda \eta. \quad (2)$$

The conventional estimate for η assumes λ is known and operates independently at each pixel:

$$\hat{\eta}^{\text{baseline}}(Y, \lambda) = \frac{Y}{\lambda}. \quad (3)$$

When λ is assumed to be constant, but actually varies in time, the conventional estimate (3) gives rise to prominent stripe artifacts like the ones shown in Figures 1b and 1d.

B. Continuous-Time Time-Resolved Measurement

The continuous-time (CT) measurement model is an idealization of an FIB microscope first introduced in [14]. It allows us to study the limits of TR measurement. Here, we imagine that we have direct detection of SEs with perfect temporal precision of each SE burst. In our model, some incident ions result in no detected SEs because $P(X_i = 0)$ is nonzero. There is no observed difference between the lack of an incident ion and an incident ion resulting in no detected SEs. Thus, the CT measurement is

$$\left\{ \tilde{M}, (\tilde{T}_1, \tilde{X}_1), (\tilde{T}_2, \tilde{X}_2), \dots, (\tilde{T}_{\tilde{M}}, \tilde{X}_{\tilde{M}}) \right\}, \quad (4)$$

where $\tilde{M}$ is the number of incident ions that result in at least one SE, $\tilde{T}_i$ is the time of the i th such incident ion, and $\tilde{X}_i$ is the corresponding SE counts. The probability of an incident ion yielding at least one SE is $1 - e^{-\eta}$, so we have $\tilde{M} \sim \text{Poisson}(\lambda(1 - e^{-\eta}))$ with PMF

$$P_{\tilde{M}}(\tilde{m}; \eta, \lambda) = \exp(-\lambda(1 - e^{-\eta})) \frac{(\lambda(1 - e^{-\eta}))^{\tilde{m}}}{\tilde{m}!}, \quad (5)$$

$m = 0, 1, \dots$. The X_i s are independent and identically distributed with PMF

$$P_{\tilde{X}_i}(j; \eta) = \frac{e^{-\eta}}{1 - e^{-\eta}} \cdot \frac{\eta^j}{j!}, \quad j = 1, 2, \dots, \quad (6)$$

which is the zero-truncation of the Poisson(η) distribution.

Several estimators of η from a CT measurement when λ is known were studied in [14]. Here, we are interested in the *assumed dose* $\tilde{\lambda}$ not necessarily being equal to the dose λ . The CT maximum likelihood (ML) estimate evaluated using the assumed dose $\tilde{\lambda}$ is the unique root of the following equation:

$$\hat{\eta}^{\text{CT}|\tilde{\lambda}} = \frac{Y}{\tilde{M} + \tilde{\lambda} e^{-\hat{\eta}^{\text{CT}|\tilde{\lambda}}}}. \quad (7)$$

C. Discrete-Time Time-Resolved Measurement

The discrete-time (DT) measurement model assumes the per-pixel dwell time t is split into n sub-acquisitions of equal duration. Each sub-acquisition then has the same distribution as a conventional measurement with dose λ/n . A key observation of this paper is that the n -length DT measurement vector

contains rich information about both the dose λ and SE yield η . Imagine short sub-acquisitions with dose λ/n small enough that observing more than one incident ion per sub-acquisition is unlikely. In this case, with large enough η , the number of sub-acquisitions where the number of observed SEs is strictly positive is roughly equal to the number of incident ions.

In this work, we will use subscript k for pixel index, so pixel k has beam current λ_k and SE yield η_k ; vectors $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ contain the values of η and λ for each of the p pixels in a sample. The vector $\mathbf{y} \in \mathbb{R}^{pn}$ gathers measurements across all pixels, with the vector of n time-resolved measurements at the k th pixel given by $\mathbf{y}_k = [y_k^{(1)}, y_k^{(2)}, \dots, y_k^{(n)}]$. Since the entries in $\mathbf{y}_k$ are independent, their joint PMF is

$$P_{\mathbf{y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) = \prod_{i=1}^n P_Y(y_k^{(i)}; \eta_k, \lambda_k/n), \quad (8)$$

where $P_Y(\cdot; \cdot, \cdot)$ is given by (1). Under assumed dose $\tilde{\lambda}$, the DT ML estimator finds the value of η_k , separately at each sample pixel, that maximizes the likelihood in (8):

$$\hat{\eta}_k^{\text{DT}|\tilde{\lambda}}(\mathbf{y}_k, \tilde{\lambda}) = \arg \max_{\eta_k} \prod_{i=1}^n P_Y(y_k^{(i)}; \eta_k, \tilde{\lambda}/n). \quad (9)$$

In this work, we use TR data to estimate *both* λ and η . When oracle knowledge of the dose λ is assumed while estimating η , the estimate is denoted $\hat{\eta}_k^{\text{DT}|\lambda}$.

III. FEASIBILITY OF JOINT ESTIMATION OF SE YIELD AND BEAM CURRENT

A single conventional measurement Y combines SE yield η and beam current λ inseparably, as suggested by (2). However, in this section, we show that using time-resolved data, joint estimation of η and λ becomes possible. In Section III-A we derive CT and DT CRBs for the joint estimation problem. In Section III-B, we discuss the challenge of joint estimation at low- η pixels.

A. Cramér–Rao Bound

The CRB provides a lower bound for the variance of an unbiased estimator of our unknown parameter $\theta = [\eta, \lambda]$. In this section we derive the CRB for joint estimation of η and λ under both CT and DT measurement models and use them to explore the feasibility and challenge of joint estimation.

1) *Continuous-Time Cramér–Rao Bound:* With a CT measurement (4), the entries of the Fisher information (FI) matrix are given by

$$[I^{\text{CT}}]_{i,j} = E \left[\left(\frac{\partial \log P_{\tilde{m}, \tilde{T}, \tilde{X}}(\tilde{m}, \tilde{t}, \tilde{x}; \eta, \lambda)}{\partial \theta_i} \right) \left(\frac{\partial \log P_{\tilde{m}, \tilde{T}, \tilde{X}}(\tilde{m}, \tilde{t}, \tilde{x}; \eta, \lambda)}{\partial \theta_j} \right) \middle| \eta, \lambda \right]. \quad (10)$$

As shown in [14, Sect. III-B], the FI about η in the CTTR measurement is

$$[I^{\text{CT}}(\eta, \lambda)]_{1,1} = \lambda \left(\frac{1}{\eta} - e^{-\eta} \right). \quad (11)$$

The FI about λ in the CTTR measurement is

$$\begin{aligned} [\mathcal{I}^{\text{CT}}]_{2,2} &\stackrel{(a)}{=} \mathbb{E} \left[\widetilde{M} \right] \mathcal{I}_{\widetilde{X}_i}(\lambda; \eta) + \mathcal{I}_{\widetilde{M}}(\lambda; \eta) \\ &\stackrel{(b)}{=} \mathcal{I}_{\widetilde{M}}(\lambda; \eta), \end{aligned} \quad (12)$$

where (a) follows from the application of the chain rule for FI, the fact that the conditional distribution of $\widetilde{T}_i$ given $\widetilde{M}$ has no dependence on η , and the independence of $\{\widetilde{M}, \widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_{\widetilde{M}}\}$; and (b) from the fact that $P_{\widetilde{X}_i}(j; \eta)$ is not a function of λ and thus $\mathcal{I}_{\widetilde{X}_i}(\lambda; \eta) = 0$. It follows that

$$\begin{aligned} [\mathcal{I}^{\text{CT}}]_{2,2} &= \mathcal{I}_{\widetilde{M}(\lambda; \eta)} = \mathbb{E} \left[\left(\frac{\partial \log P_{\widetilde{M}}(\widetilde{M}; \lambda, \eta)}{\partial \lambda} \right)^2; \lambda \right] \\ &= \sum_{j=0}^{\infty} \left(-(1 - e^{-\eta}) + \frac{j}{\lambda} \right)^2 \frac{e^{-\lambda(1-e^{-\eta})} [\lambda(1-e^{-\eta})]^j}{j!} \\ &= \frac{1 - e^{-\eta}}{\lambda}. \end{aligned} \quad (13)$$

The cross terms in the the FI matrix are

$$\begin{aligned} [\mathcal{I}^{\text{CT}}]_{1,2} &= [\mathcal{I}^{\text{CT}}]_{2,1} \\ &= \mathbb{E} \left[\left(\frac{\partial \log P_{\widetilde{M}}(\widetilde{M}; \lambda, \eta)}{\partial \lambda} \right) \left(\frac{\partial \log P_{\widetilde{M}}(\widetilde{M}; \lambda, \eta)}{\partial \eta} \right) \right] \\ &= \sum_{j=0}^{\infty} \left(-(1 - e^{-\eta}) + \frac{j}{\lambda} \right) \left(\frac{e^{-\eta}}{1 - e^{-\eta}} j - \lambda e^{-\eta} \right) \\ &\quad \cdot \frac{e^{-\lambda(1-e^{-\eta})} [\lambda(1-e^{-\eta})]^j}{j!} \\ &= e^{-\eta}. \end{aligned} \quad (14)$$

Thus, the full CT FI matrix is

$$\mathcal{I}^{\text{CT}} = \begin{bmatrix} \lambda \left(\frac{1}{\eta} - e^{-\eta} \right) & e^{-\eta} \\ e^{-\eta} & \frac{1}{\lambda} (1 - e^{-\eta}) \end{bmatrix}. \quad (15)$$

When one parameter is to be estimated and the other is given, the CT CRB is given by

$$\sigma_{\eta|\lambda}^2 \geq \text{CRB}^{\text{CT}}(\eta|\lambda) = [\mathcal{I}^{\text{CT}}]_{1,1}^{-1} = \frac{1}{\lambda} \left(\frac{1}{\eta} - e^{-\eta} \right)^{-1}, \quad (16a)$$

$$\sigma_{\lambda|\eta}^2 \geq \text{CRB}^{\text{CT}}(\lambda|\eta) = [\mathcal{I}^{\text{CT}}]_{2,2}^{-1} = \frac{\lambda}{1 - e^{-\eta}}. \quad (16b)$$

When both η and λ are unknown, the CT CRB for each parameter is computed by inverting $\mathcal{I}_{\text{CT}}$:

$$\sigma_{\eta}^2 \geq \text{CRB}^{\text{CT}}(\eta) = [\mathcal{I}^{\text{CT}}]_{1,1}^{-1} = \frac{1 - e^{-\eta}}{\frac{\lambda}{\eta}(1 - e^{-\eta}) - \lambda e^{-\eta}}, \quad (17a)$$

$$\sigma_{\lambda}^2 \geq \text{CRB}^{\text{CT}}(\lambda) = [\mathcal{I}^{\text{CT}}]_{2,2}^{-1} = \frac{\lambda(1 - \eta e^{-\eta})}{(1 - e^{-\eta}) - \eta e^{-\eta}}. \quad (17b)$$

Note that the CRBs when both parameters are unknown (17) may be written in terms of the CRBs for that same parameter when the other parameter is known (16):

$$\text{CRB}_{\text{CT}}(\eta) = \alpha(\eta) \text{CRB}_{\text{CT}}(\eta|\lambda), \quad (18a)$$

$$\text{CRB}_{\text{CT}}(\lambda) = \alpha(\eta) \text{CRB}_{\text{CT}}(\lambda|\eta), \quad (18b)$$

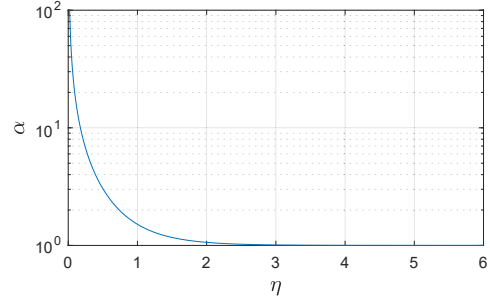


Fig. 2: Factor from (19) by which the CRB for joint estimation of η and λ exceeds that of estimating one parameter with the other given.

where

$$\alpha(\eta) = \frac{1 - (1 + \eta)e^{-\eta} + \eta e^{-2\eta}}{1 - (1 + \eta)e^{-\eta}}. \quad (19)$$

The factor $\alpha(\eta)$, which is plotted in Figure 2, represents the added challenge of the joint estimation problem compared to estimating one parameter given the other. When $\eta \geq 2$, as is typical for FIB microscopy, $\alpha(\eta) \approx 1$, and furthermore $\lim_{\eta \rightarrow \infty} \alpha(\eta) = 1$; i.e., asymptotically, jointly estimating both parameters is no more challenging. When η is low, $\widetilde{M}$ becomes a less suitable proxy for, and contains less information about, the number of incident ions M .

2) *Discrete-Time Cramér–Rao Bound*: The Fisher information matrix about unknown parameter θ in n time resolved measurements, each with a per sub-acquisition dose of λ/n is given by

$$\begin{aligned} [\mathcal{I}^{\text{DT}}(\eta, \lambda; n)]_{i,j} &= n \mathbb{E} \left[\left(\frac{\partial \log P_Y(y; \eta, \frac{\lambda}{n})}{\partial \theta_i} \right) \left(\frac{\partial \log P_Y(y; \eta, \frac{\lambda}{n})}{\partial \theta_j} \right) \middle| \eta, \lambda, n \right] \\ &= n \sum_{y=0}^{\infty} \left(\frac{\partial \log P_Y(y; \eta, \lambda/n)}{\partial \theta_i} \right) \\ &\quad \left(\frac{\partial \log P_Y(y; \eta, \lambda/n)}{\partial \theta_j} \right) P_Y(y; \eta, \lambda/n). \end{aligned} \quad (20)$$

As derived in [13],

$$\frac{\partial \log P_Y(y; \eta, \lambda/n)}{\partial \eta} = \frac{y}{\eta} - \frac{P_Y(y+1; \eta, \lambda/n)}{P_Y(y; \eta, \lambda/n)} \frac{y+1}{\eta}, \quad (21)$$

and similarly, the derivative with respect to λ is

$$\frac{\partial \log P_Y(y; \eta, \lambda/n)}{\partial \lambda} = -\frac{1}{n} + \frac{P_Y(y+1; \eta, \lambda/n)}{P_Y(y; \eta, \lambda/n)} \frac{y+1}{\eta} \frac{1}{\lambda}. \quad (22)$$

Substituting (21) and (22) into (20) and truncating the series appropriately enables numerical approximation of the FI matrix. CRBs are then analogous to (16) and (17).

B. The Challenge of Low SE Yield Samples

In Section III-A, we showed that as η gets smaller, jointly estimating η and λ becomes increasingly more difficult than estimation of η or λ alone. Figure 3 further illustrates the challenge of lower η samples. Here we plot the normalized continuous-time CRBs for $\hat{\eta}$ and $\hat{\lambda}$ as functions of the total

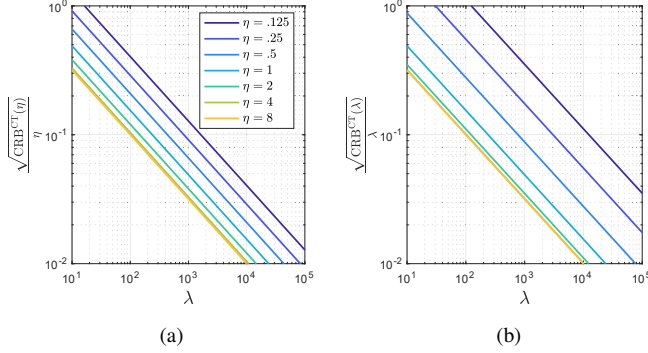


Fig. 3: Normalized CRBs as functions of λ for several values of η . As η decreases, the dose λ required to achieve any fixed desired relative error increases.

dose λ . Different colored curves denote different η values. Note that the intersection of each curve with a horizontal line indicates the dose λ required to achieve the corresponding normalized CRB. For example, in order for the normalized standard deviation to be lower bounded by 10^{-1} , a dose on the order of 10^2 is required when $\eta = 4$; when $\eta = 0.25$, a dose of nearly 10^3 is required.

In Section IV, we study joint estimation of η and λ at a *single* pixel. Although we show that this is possible, as demonstrated in Figure 3, even pixels with large η require a large dose to form high-quality estimates. In Sections V and VI, we exploit *inter-pixel* correlations of both η and λ to enable high-quality joint estimation at moderate doses.

IV. JOINT ESTIMATION AT A SINGLE PIXEL

In this section, we derive CT and DT ML estimates for η and λ using only measurements acquired at a *single* pixel. In Section IV-C, we evaluate the performance of these estimators and compare to the CRBs derived in Section III-A.

A. Continuous-Time Time-Resolved ML Estimation

Recall the continuous-time measurement model in (4). The $\tilde{X}_i$ variables are independent, so using (6), the joint distribution of the SE count vector (conditioned on $\tilde{M} = \tilde{m}$) is

$$\begin{aligned} \prod_{i=1}^{\tilde{m}} P_{\tilde{X}_i}(j_i; \eta) &= \left(\frac{e^{-\eta}}{1 - e^{-\eta}} \right)^{\tilde{m}} \frac{\eta^{j_1 + j_2 + \dots + j_{\tilde{m}}}}{j_1! j_2! \dots j_{\tilde{m}}!} \\ &= c \left(\frac{e^{-\eta}}{1 - e^{-\eta}} \right)^{\tilde{m}} \eta^y, \end{aligned} \quad (23)$$

where the simplification comes from identifying the sum as the total SE count y and replacing the product of factorials with an unspecified constant because this is immaterial to estimation of η and λ . Combining (5) and (23), the relevant likelihood is

$$\begin{aligned} P_{\tilde{M}, Y}(\tilde{m}, y | \eta, \lambda) &= c \exp(-\lambda(1 - e^{-\eta})) \frac{(\lambda(1 - e^{-\eta}))^{\tilde{m}}}{\tilde{m}!} \left(\frac{e^{-\eta}}{1 - e^{-\eta}} \right)^{\tilde{m}} \eta^y \\ &= c \exp(-\lambda(1 - e^{-\eta})) \lambda^{\tilde{m}} e^{-\eta \tilde{m}} \eta^y, \end{aligned} \quad (24)$$

where $\tilde{m}!$ has been absorbed into the constant c because this is immaterial to estimation of η and λ . Omitting the constant,

$$-\log P_{\tilde{M}, Y}(\tilde{m}, y | \eta, \lambda) = \lambda(1 - e^{-\eta}) - \tilde{m} \log \lambda + \eta \tilde{m} - y \log \eta. \quad (25)$$

Taking derivatives of $-\log P_{\tilde{M}, Y}(\tilde{m}, y | \eta, \lambda)$ gives

$$\frac{\partial}{\partial \lambda} [\sim] = (1 - e^{-\eta}) - \frac{\tilde{m}}{\lambda}, \quad (26)$$

$$\frac{\partial}{\partial \eta} [\sim] = \lambda e^{-\eta} + \tilde{m} - \frac{y}{\eta}. \quad (27)$$

Setting these to zero to find the joint ML estimate gives that $\hat{\eta}^{\text{CT}}$ is the root of

$$\frac{\eta}{1 - \exp(-\eta)} = \frac{y}{\tilde{m}}, \quad (28)$$

which then can be substituted to give

$$\hat{\lambda}^{\text{CT}} = \frac{\tilde{m}}{1 - \exp(-\hat{\eta}^{\text{CT}})}. \quad (29)$$

These values can be justified heuristically without the ML property. Since $\tilde{m}/(1 - \exp(-\eta))$ is a good proxy for the number of incident ions, (28) sets $\hat{\eta}^{\text{CT}}$ to be the number of detected SEs y divided by this estimate for the number of incident ions. In [14], this was called the *continuous-time Lambert quotient mode estimator*, and it was shown to differ from the ML estimate of η with λ known.

Note that when at most a single SE is observed in response to each of the incident ions (i.e., $\tilde{X}_i = 1$ for all i), we have $y = \tilde{m}$, so the right side of (28) equals 1. The left side of (28) approaches 1 as η approaches 0; thus, we assign $\hat{\eta}^{\text{CT}} = 0$, and substituting in (29) gives $\hat{\lambda}^{\text{CT}} = \infty$. We address this singularity by placing a reasonable upper bound λ_{max} on our estimate for λ . The smallest nonzero estimate we can obtain for η is then $1/\lambda_{\text{max}}$. Requiring a large dose to be able to accurately estimate a small value of η at a single pixel is consistent with the normalized CRBs in Figure 3a. This limitation is one motivation for our use of inter-pixel correlations in Sections V and VI.

B. Discrete-Time Time-Resolved ML Estimation

At the k th pixel, we acquire a vector $\mathbf{y}_k$ of n time-resolved measurements with joint PMF given in (8). The corresponding joint ML estimate is

$$\begin{aligned} (\hat{\eta}_k^{\text{DT}}, \hat{\lambda}_k^{\text{DT}}) &= \arg \max_{\eta_k, \lambda_k} P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) \\ &= \arg \min_{\eta_k, \lambda_k} [-\log P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k)]. \end{aligned} \quad (30)$$

The objective function in (30) is a sum of n terms, each a logarithm of the Neyman Type A PMF in (1). While difficult to work with analytically, since the decision variable is only two-dimensional and the objective function is smooth, numerical evaluation of (30) is not difficult. The numerical experiments in following sections use gradient descent methods based on derivatives (58a) and (61a) derived in the appendix.

Similar to the CT case in Section IV-A, observing at most a single SE per sub-acquisition creates a singularity whereby (30) gives $(\hat{\eta}_k^{\text{DT}}, \hat{\lambda}_k^{\text{DT}}) = (0, \infty)$. In practice, we can again place a reasonable upper bound λ_{max} on our λ estimate and then estimate η accordingly.

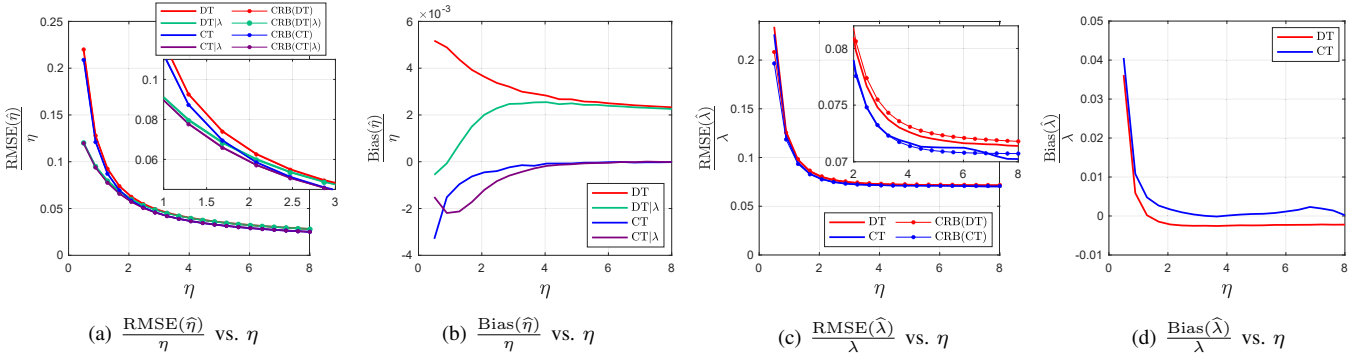


Fig. 4: Normalized RMSE and bias as functions of η for single-pixel estimators with $\lambda = 200$ and, in discrete-time cases, $\lambda/n = 0.1$. In panels (a) and (c), the normalized square roots of the Cramér–Rao bounds are plotted for reference.

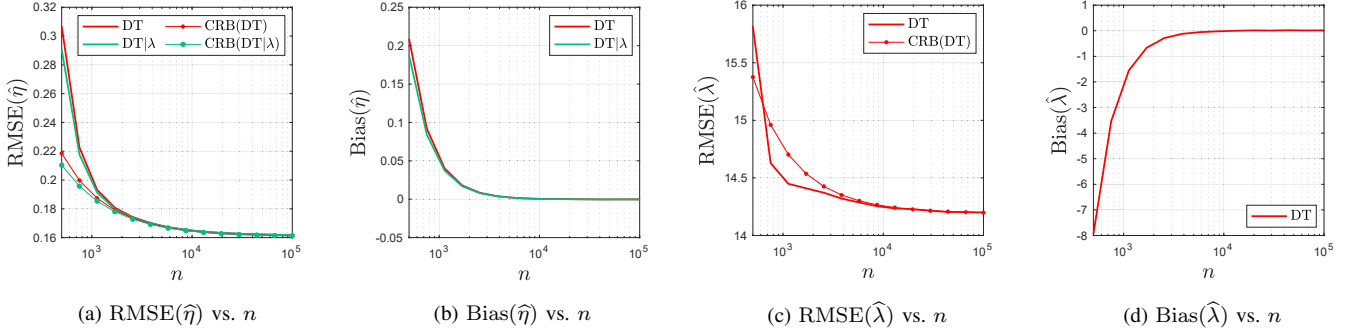


Fig. 5: RMSE and bias as functions of n for single-pixel estimators with $\lambda = 200$ and $\eta = 5$.

C. Estimator Performance

In Figure 4, we plot the root mean-squared error (RMSE) and bias for our estimators, both normalized by the true parameter value, as functions of η . We compare existing methods for estimating η , which require knowledge of the dose, with our new methods for jointly estimating η and λ :

- CT| λ : $\hat{\eta}_k^{\text{CT}|\lambda}$ is the CT estimate in (7), evaluated using oracular knowledge of the dose ($\tilde{\lambda} = \lambda$).
- DT| λ : $\hat{\eta}_k^{\text{DT}|\lambda}$ is the DT estimate in (9), evaluated using oracular knowledge of the dose ($\tilde{\lambda} = \lambda$).
- CT: $(\hat{\eta}_k^{\text{CT}}, \hat{\lambda}_k^{\text{CT}})$ is the CT joint estimate from (28) and (29)
- DT: $(\hat{\eta}_k^{\text{DT}}, \hat{\lambda}_k^{\text{DT}})$ is the DT joint estimate from (30).

We also plot the normalized square roots of the corresponding CRBs: CRB(CT) of (17), CRB(CT| λ) of (16a), as well as CRB(DT) and CRB(DT| λ) computed as described in Section III-A2. As predicted by the CRBs derived in Section III-A, our joint estimators and the estimators of η when λ is given achieve similar performance at higher η . In fact, Figures 4a and 4c both show a very close match of all estimators to their corresponding CRBs, with the (biased) DT estimate of λ slightly outperforming the CRB. With $\lambda = 200$ as used here, the normalized RMSEs of both $\hat{\eta}$ and $\hat{\lambda}$ get very large as η gets smaller, especially when λ is not given, motivating our use of inter-pixel correlations in Sections V and VI. The RMSE and bias of $\hat{\eta}$ and $\hat{\lambda}$ are generally smaller for CT methods, with very small bias of $\hat{\eta}$ and $\hat{\lambda}$ for all estimators.

Figure 5 shows estimator RMSE (along with the square root of the CRBs) and bias as functions of the number of sub-

acquisitions n for fixed total dose $\lambda = 200$ and SE yield $\eta = 5$. Note that when n gets large, DT performance converges to the CT asymptote. At $\lambda/n = 0.1$, a value attainable by current hardware and used in the DT experiments that follow in Section V-G, DT estimator RMSE is close to the CT limit. Figures 5a and 5b show that joint estimation and estimation of η given λ are similarly difficult when n is sufficiently large. In Figure 5a we observe that at larger n , $\text{RMSE}(\hat{\eta}^{\text{DT}})$ approaches the CRB. The RMSE of $\hat{\lambda}^{\text{DT}}$ dips below the CRB (Figure 5c) at certain lower values of n , which may be explained by the non-negligible bias shown in Figure 5d.

V. EXPLOITING A SMOOTHLY VARYING BEAM CURRENT

Section IV demonstrated that joint estimation of η and λ is possible at a single pixel through time-resolved measurement. However, as shown in Figure 3 and discussed in Section IV-A, high-fidelity estimates may require a large dose, especially at low- η pixels. In this section, we use a simple model for smoothly varying beam current—meant to be representative of electron and helium ion beams—to form high-quality estimates of both η and λ at moderate doses. To meet a variety of use cases, we propose both causal and non-causal algorithms, each with and without total variation (TV) regularization on η . Note that TV regularization is just one effective example; our techniques may be extended to apply other regularization techniques. In the interests of brevity and relevance to contemporary instruments, we consider only discrete-time measurements.

A. An Autoregressive Model for Beam Current

We model beam current as a first-order Gaussian autoregressive process:

$$\lambda_k = x_k + a\lambda_{k-1} + c, \quad (31)$$

where a is the correlation coefficient for neighboring pixels in a row and all $x_k \sim \mathcal{N}(0, \sigma_x^2)$ variables are independent. The mean and variance of the beam current are

$$\bar{\lambda} = \mathbb{E}[\lambda] = \frac{c}{1-a} \quad \text{and} \quad \sigma_\lambda^2 = \frac{\sigma_x^2}{1-a^2}. \quad (32)$$

This describes an incident beam with slow, unknown variations about $\bar{\lambda}$, which may be the intended beam current setting.

B. Causal Estimation

We seek *causal* estimates, $\hat{\eta}_k^C$ and $\hat{\lambda}_k^C$, of η_k and λ_k at the k th pixel, given only the measurement vector at that pixel $\mathbf{Y}_k = \mathbf{y}_k$ and λ_{k-1} . Although λ_{k-1} is not perfectly known, we assume it is approximately equal to our estimate of the beam current formed at the last pixel: $\lambda_{k-1} \approx \hat{\lambda}_{k-1}^C$. Inspired by (31), we use the prior

$$\lambda_k | \lambda_{k-1} \sim \mathcal{N}(a\hat{\lambda}_{k-1}^C + c, \sigma_x^2) \quad (33)$$

and formulate a MAP estimate:

$$\begin{aligned} (\hat{\eta}_k^C, \hat{\lambda}_k^C) &= \arg \max_{\eta_k, \lambda_k} f(\eta_k, \lambda_k | \mathbf{y}_k) \\ &= \arg \max_{\eta_k, \lambda_k} P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) f(\lambda_k | \lambda_{k-1}) \\ &= \arg \min_{\eta_k, \lambda_k} \left[-\log P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) \right. \\ &\quad \left. + \frac{1}{2\sigma_x^2} (\lambda_k - (a\hat{\lambda}_{k-1}^C + c))^2 \right], \end{aligned} \quad (34)$$

where $P_{\mathbf{Y}_k}(\cdot; \cdot, \cdot)$ is the joint PMF given in (8). In practice, we introduce a tuning parameter β_C :

$$\begin{aligned} (\hat{\eta}_k^C, \hat{\lambda}_k^C) &= \arg \min_{\eta_k, \lambda_k} \left[-\log P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) \right. \\ &\quad \left. + \beta_C (\lambda_k - (a\hat{\lambda}_{k-1}^C + c))^2 \right]. \end{aligned} \quad (35)$$

At each new pixel, (35) is solved using gradient descent. At the first pixel, indexed by $k = 1$, we solve (35) using $\hat{\lambda}_0^C = \bar{\lambda}$.

C. Causal Estimation with Total Variation Regularization

Total variation regularization on η may be added to the cost function in (35) to exploit the fact that microscopy images are often piecewise smooth. Our new TV-regularized causal estimate of η (and the corresponding estimate of λ) minimize the following cost function:

$$\begin{aligned} (\hat{\eta}_k^{\text{CTV}}, \hat{\lambda}_k^{\text{CTV}}) &= \arg \min_{\eta_k, \lambda_k} \left[-\log P_{\mathbf{Y}_k}(\mathbf{y}_k; \eta_k, \lambda_k) \right. \\ &\quad \left. + \beta_C (\lambda_k - (a\lambda_{k-1} + c))^2 + g_{\text{TV}}(\eta_k) \right], \end{aligned} \quad (36)$$

where $g_{\text{TV}}(\eta_k)$ is a TV cost term. In the causal, raster-scanned scenario, the neighboring pixels that have already been visited

are those to the left and above the current pixel. Thus, our TV cost term is given by

$$g_{\text{TV}}(\eta) = \beta_h |\eta - \eta_h| + \beta_v |\eta - \eta_v|, \quad (37)$$

where η_v and η_h are η values already estimated (and assumed known) at the vertically and horizontally adjacent pixels. Parameters β_h and β_v may be tuned to promote more horizontal or vertical smoothness.

We solve (36) using proximal gradient methods. The proximal operator for the term in (37) is

$$\text{prox}_{g_{\text{TV}}}(x) = \arg \min_{\alpha} \frac{1}{2} \|x - \alpha\|_2^2 + \beta_h |\alpha - \eta_h| + \beta_v |\alpha - \eta_v|. \quad (38)$$

When $\eta_h < \eta_v$ holds, the minimization in (38) gives

$$\text{prox}_{g_{\text{TV}}}(x) = \begin{cases} x + \beta_h + \beta_v, & \text{if } x < \eta_h - \beta_h - \beta_v; \\ x - \beta_h - \beta_v, & \text{if } x > \eta_v + \beta_h + \beta_v; \\ x - \beta_h + \beta_v, & \text{if } \eta_h + \beta_h - \beta_v < x \\ & < \eta_v + \beta_h - \beta_v; \\ \eta_h, & \text{if } |x - \eta_h + \beta_v| \leq \beta_h; \\ \eta_v, & \text{if } |x - \eta_v - \beta_v| \leq \beta_v, \end{cases} \quad (39)$$

which, when $\beta_h = \beta_v = \beta_{\text{CTV}}$,² reduces to

$$\text{prox}_{g_{\text{TV}}}(x) = \begin{cases} x + 2\beta_{\text{CTV}}, & \text{if } x < \eta_h - 2\beta_{\text{CTV}}; \\ x - 2\beta_{\text{CTV}}, & \text{if } x > \eta_v + 2\beta_{\text{CTV}}; \\ x, & \text{if } \eta_h < x < \eta_v; \\ \eta_h, & \text{if } \eta_h - 2\beta_{\text{CTV}} < x < \eta_h; \\ \eta_v, & \text{if } \eta_v \leq x \leq \eta_v + 2\beta_{\text{CTV}}. \end{cases} \quad (40)$$

The case of $\eta_h > \eta_v$ is similar. Equation (36) is solved at each pixel, with $\beta_v = 0$ at all pixels in the first row of the image and $\beta_h = 0$ for the first column of the image.

D. Non-Causal Estimation

The non-causal estimation algorithm operates on the *entire* measurement vector $\mathbf{y}$ and estimates η and λ simultaneously at all pixels. This formulation allows us to leverage stronger priors on λ , as well as on η as we show later in Section V-E. Given (31), λ is jointly Gaussian,

$$\lambda \sim \mathcal{N}(\bar{\lambda}\mathbf{1}, \Sigma), \quad \Sigma_{i,j} = \frac{\sigma_x^2}{1-a^2} a^{|i-j|}, \quad (41)$$

where $\mathbf{1} \in \mathbb{R}^p$ is a vector of ones. Since all sub-acquisitions and pixels are conditionally independent, the joint PMF of the *entire* measurement vector $\mathbf{y}$ is

$$P_{\mathbf{Y}}(\mathbf{y} | \eta, \lambda) = \prod_{k=1}^p \prod_{i=1}^n P_Y(y_k^{(i)}; \eta_k, \lambda_k/n), \quad (42)$$

²In this paper, we evaluate our algorithm with $\beta_h = \beta_v = \beta_{\text{CTV}}$. However promoting more similarity between pixels that are vertically adjacent (i.e., $\beta_v > \beta_h$) might be useful to mitigate horizontal stripe artifacts.

where $P_Y(\cdot; \cdot, \cdot)$ is the PMF in (1). The MAP estimate for $(\eta, \lambda, \bar{\lambda})$ is given by

$$\begin{aligned} (\hat{\eta}^{\text{NC}}, \hat{\lambda}^{\text{NC}}, \hat{\bar{\lambda}}) &= \arg \max_{\eta, \lambda, \bar{\lambda}} P_Y(\mathbf{y} | \eta, \lambda) f(\lambda | \bar{\lambda}) \\ &= \arg \min_{\eta, \lambda, \bar{\lambda}} \left[-\log P_Y(\mathbf{y} | \eta, \lambda) + \frac{1}{2} (\lambda - \bar{\lambda} \mathbf{1})^\top \Sigma^{-1} (\lambda - \bar{\lambda} \mathbf{1}) \right]. \end{aligned} \quad (43)$$

As in (35), we introduce a tuning parameter β_{NC} to allow additional regularization:

$$\begin{aligned} (\hat{\eta}^{\text{NC}}, \hat{\lambda}^{\text{NC}}, \hat{\bar{\lambda}}) &= \arg \min_{\eta, \lambda, \bar{\lambda}} \left[-\log P_Y(\mathbf{y} | \eta, \lambda) \right. \\ &\quad \left. + \beta_{\text{NC}} (\lambda - \bar{\lambda} \mathbf{1})^\top \Sigma^{-1} (\lambda - \bar{\lambda} \mathbf{1}) \right]. \end{aligned} \quad (44)$$

Note that this formulation does not require knowledge of the mean beam current but rather estimates $\bar{\lambda}$ in addition to η and λ . This cost function is differentiable and thus we solve the minimization using gradient descent methods. The derivatives of the first term in (44) are derived in the appendix; the derivative of the second term with respect to λ is proportional to $\Sigma^{-1}(\lambda - \bar{\lambda} \mathbf{1})$. To avoid storing a prohibitively large matrix, and because Σ is approximately circulant, we perform multiplication by Σ^{-1} in the frequency domain using the fast Fourier transform.

E. Non-Causal Estimation with Total Variation Regularization

As with our causal estimate, TV regularization may be added to (44) to promote piecewise smooth estimates of η :

$$\begin{aligned} (\hat{\eta}^{\text{NCTV}}, \hat{\lambda}^{\text{NCTV}}, \hat{\bar{\lambda}}) &= \arg \min_{\eta, \lambda, \bar{\lambda}} \left[-\log P_Y(\mathbf{y} | \eta, \lambda) \right. \\ &\quad \left. + \beta_{\text{NC}} (\lambda - \bar{\lambda} \mathbf{1})^\top \Sigma^{-1} (\lambda - \bar{\lambda} \mathbf{1}) + \beta_{\text{NCTV}} \|\eta\|_{\text{TV}} \right], \end{aligned} \quad (45)$$

where $\|\eta\|_{\text{TV}}$ is given by

$$\|\eta\|_{\text{TV}} = \sum_{i,j} \sqrt{|\eta_{i+1,j} - \eta_{i,j}|^2 + |\eta_{i,j+1} - \eta_{i,j}|^2} \quad (46)$$

and β_{NCTV} is a tuning parameter. Equation (45) is solved with proximal gradient methods using the proximal operator for (46) given in [24].

F. Operational Considerations

Although all of our proposed algorithms jointly estimate η and λ , different assumptions are made about the parameters a , σ_x^2 and $\bar{\lambda}$ in (31). Note that both causal and non-causal algorithms assume knowledge of the correlation a between pixels. Due to the tuning parameters β_C and β_{NC} , no assumption is made about σ_x^2 . In practice, algorithm performance was found to not depend heavily on ideal choices of a , β_C , or β_{NC} , as we will show in Section V-G, Figures 10 and 11. While our causal algorithms do require knowledge of the mean beam current $\bar{\lambda}$, the non-causal algorithms estimate $\bar{\lambda}$ in addition to λ and η . When causal operation is warranted but the mean beam current is not known, the non-causal algorithm could be run periodically to provide $\bar{\lambda}$.

G. Simulated Microscopy Results

1) *Data Generation*: We evaluate the multi-pixel algorithms proposed in this section on synthetic HIM and SEM data. Measurements for these two examples were generated using existing micrographs as ground truth images.³ Compared to SEM, HIM has higher SE yield and can thus produce high-quality images at lower doses. To be representative of HIM, we scale the ground truth to $\eta \in [2, 8]$ and use mean dose $\bar{\lambda} = 20$ [25]; for SEM, we use $\eta \in [0.1, 1]$ and $\bar{\lambda} = 200$ [26]. Beam current time series were produced according to the Gaussian first-order autoregressive model in (31). In both test examples, the correlation coefficient for neighboring pixels in a row is $a = 0.999$ and the coefficient of variation is $\sigma_\lambda/\bar{\lambda} = 0.2$. Data is generated pseudorandomly at each pixel following the separable joint PMF in (8), where in each case the nominal sub-acquisition doses is 0.1 ($n = 200$ for HIM and $n = 2000$ for SEM).

2) *Methods*: We compare nine methods for estimating η , some of which also generate an estimate of λ :

- baseline: $\hat{\eta}^{\text{baseline}}$ is the pixel-wise evaluation of (3) independently at each pixel using the nominal dose $\bar{\lambda}$.
- frequency-domain filter (FDF) [5]: Compute the 2D discrete Fourier transform of $\hat{\eta}^{\text{baseline}}$. Let q and u be the horizontal and vertical frequency indexes. Coefficients that satisfy both $|q| \leq w$ and $|u| > h$ are nulled before applying the inverse transform to yield $\hat{\eta}^{\text{FT}}$.
- DT| λ : $\hat{\eta}^{\text{DT}|\lambda}$ is the pixel-wise ML estimate (9) computed with true beam current λ (provided by an oracle).
- DT| $\bar{\lambda}$: $\hat{\eta}^{\text{DT}|\bar{\lambda}}$ is the pixel-wise ML estimate (9) computed with nominal beam current $\bar{\lambda}$.
- linear filter: $(\hat{\eta}^{\text{LF}}, \hat{\lambda}^{\text{LF}})$ is the joint estimate computed with the method of [27].
- causal: $(\hat{\eta}^{\text{C}}, \hat{\lambda}^{\text{C}})$ is the joint estimate from (35).
- non-causal: $(\hat{\eta}^{\text{NC}}, \hat{\lambda}^{\text{NC}})$ is the joint estimate from (44). (This also produces an estimate of the mean beam current $\bar{\lambda}$.)
- causal with TV: $(\hat{\eta}^{\text{CTV}}, \hat{\lambda}^{\text{CTV}})$ is the joint estimate from (36).
- non-causal with TV: $(\hat{\eta}^{\text{NCTV}}, \hat{\lambda}^{\text{NCTV}})$ is the joint estimate from (45). (This also produces an estimate of the mean beam current $\bar{\lambda}$.)

Parameters in the frequency-domain filter, linear filter, and methods from Sections V-B to V-E are tuned to minimize RMSE.

3) *Results*: Figures 6 and 7 show estimated micrographs $\hat{\eta}$ using all the methods on each setting. The inset images show error $\hat{\eta} - \eta$ for portions of the micrographs. In both HIM and SEM examples, stripe artifacts are more prominent in $\hat{\eta}^{\text{baseline}}$ (Figures 6f and 7f) than in the TR reconstruction $\hat{\eta}^{\text{DT}|\bar{\lambda}}$ (Figures 6g and 7g), although they both assume the same knowledge of λ . In the higher η FIB example, $\hat{\eta}^{\text{DT}|\bar{\lambda}}$ even outperforms $\hat{\eta}^{\text{FT}}$, the frequency-domain filtered version of $\hat{\eta}^{\text{baseline}}$. This effect may be attributed to the natural robustness of TR methods to unknown beam current [15], [16]. Our oracle TR method $\hat{\eta}^{\text{DT}|\lambda}$ (Figures 6b and 7b), which assumes

³All ground truth images in this work are from the ThermoFisher Scientific database: <https://www.fei.com/image-gallery/>

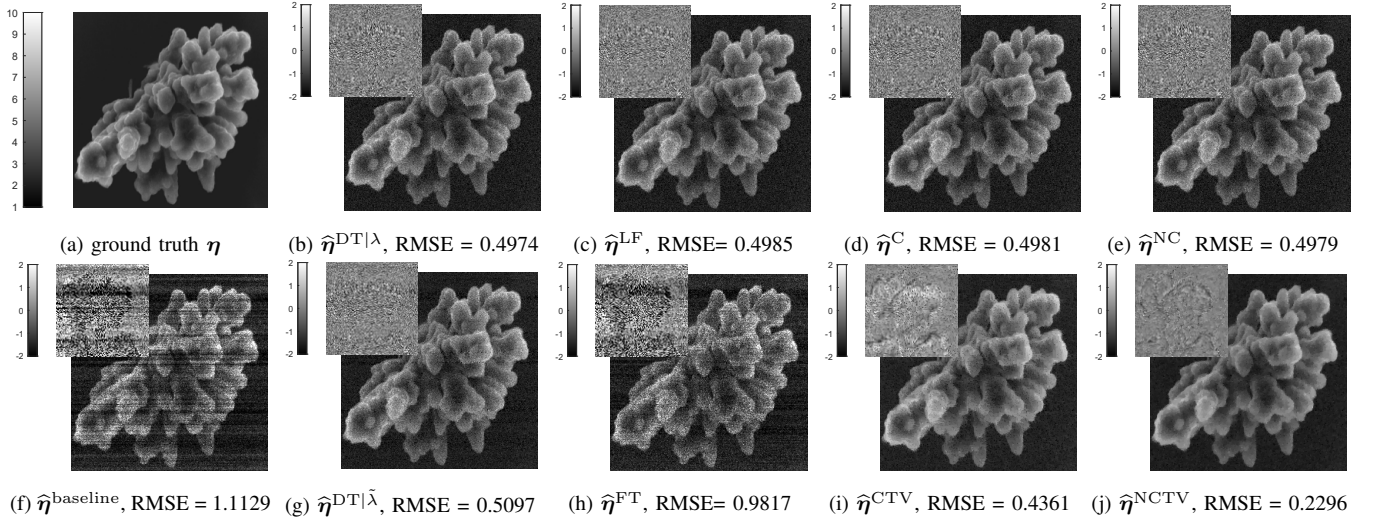


Fig. 6: HIM example with ground truth $\eta \in [2, 8]$, mean dose $\bar{\lambda} = 20$ and nominal sub-acquisition dose $\bar{\lambda}/n = 0.1$. The actual dose λ is a Gaussian autoregressive process with correlation coefficient of 0.999 for neighboring pixels in a row and coefficient of variation $\sigma_{\lambda}/\bar{\lambda} = 0.2$. All micrograph images are on the same scale shown in (a), chosen so that no more than 2% of pixels are saturated in any given image. Inset images show error $\hat{\eta} - \eta$ for a subset of the image taken from the top right corner. Tuning parameters are: $\beta_{NC} = 200$, $\beta_{NCTV} = 1$, $\beta_C = 10$, and $\beta_{CTV} = 1e-4$. The non-causal estimator found $\hat{\lambda} = 20.41$ and the non-causal estimator with TV regularization found $\hat{\lambda} = 20.29$; the beam current empirical mean was $\frac{1}{p} \sum_{k=1}^p \lambda_k = 20.34$.

Method	HIM Example		SEM Example	
	RMSE($\hat{\eta}$)	RMSE($\hat{\lambda}$)	RMSE($\hat{\eta}$)	RMSE($\hat{\lambda}$)
Baseline	1.1129	—	9.72e-2	—
FDF [5]	0.9817	—	6.93e-2	—
DT λ	0.4974	—	5.54e-2	—
DT $\hat{\lambda}$ [13]	0.5097	—	8.87e-2	—
Linear filter [27]	0.4985	1.0020	6.63e-2	17.8849
Causal	0.4984	0.9416	5.83e-2	9.4937
Non-causal	0.4979	0.6765	5.68e-2	6.6761
Causal with TV	0.4361	1.0215	5.54e-2	11.7524
Non-causal with TV	0.2298	0.6681	3.21e-2	5.8292

TABLE I: RMSE results by method for the HIM example in Figure 6 and the SEM example in Figure 7. For the frequency-domain filtering method, filter parameters were $w = 1$ and $h = 5$ for the HIM example and $w = 1$ and $h = 1$ for the SEM example. Our new joint estimation methods without TV regularization approach the performance of oracle estimator $\hat{\eta}^{DT|\lambda}$. When TV regularization is added, our causal and non-causal estimators outperform $\hat{\eta}^{DT|\lambda}$.

perfect knowledge of the true dose at every pixel, exhibits no discernible striping. Without TV regularization, our causal joint estimates $\hat{\eta}^C$ (Figures 6d and 7d) and non-causal joint estimates $\hat{\eta}^{NC}$ (Figures 6e and 7e) exhibit lower RMSE, closer to our benchmark $\hat{\eta}^{DT|\lambda}$, with slightly more improvement seen with the non-causal version. Given TV regularization on η , both algorithms meet or exceed the performance of $\hat{\eta}^{DT|\lambda}$, which does not employ any spatial regularization. In fact, in both examples, $\hat{\eta}^{NCTV}$ (Figures 6j and 7j) outperforms the benchmark $\hat{\eta}^{DT|\lambda}$ approximately by a factor of 2. Table I summarizes the RMSE results of all η estimation methods for both HIM and SEM examples.

Beam current estimates $\hat{\lambda}^C$ and $\hat{\lambda}^{NC}$ are shown in Figure 8 for the HIM example and in Figure 9 for the SEM example. Both estimates closely match the true beam current. The causal

estimate $\hat{\lambda}^C$ has higher RMSE with a slight lag and more higher frequency noise. All of our joint estimators outperform the initial joint estimator $\hat{\eta}^{LF}$ introduced in [27]. The RMSE results of different λ estimates, for both HIM and SEM examples, are summarized in Table I.

As previously noted, all of our joint reconstruction algorithms include a tuning parameter (β_C or β_{NC} to control regularization on λ) and assume knowledge of correlation a . In Figure 10, we show that performance of our causal estimation algorithm is not heavily dependent on ideal choices of either of these two parameters. Using the data from Figure 6, Figures 10a and 10b form the joint estimate ($\hat{\eta}^C$, $\hat{\lambda}^C$) for different values of β_C , keeping the correlation fixed at $a = 0.999$. Figures 10c and 10d fix $\beta_C = 10$ and vary the assumed value of a . We note that different values of β_C and a have a negligible effect on the RMSE of $\hat{\eta}^C$. Although the value of β_C and a does effect the RMSE of $\hat{\lambda}^C$, the RMSE remains relatively small across a wide range of values, with larger values of β_C and a (i.e., promoting more smoothing of λ) as safer choices when the true parameter value is not known. In Figure 11, we show similar trends for our non-causal algorithms.

VI. EXPLOITING A DISCRETE MARKOV BEAM CURRENT

In this section, we demonstrate joint estimation when beam current flips back and forth between two values, as a simple model for a neon beam microscope. This model is inspired by the observation in [6] that adhesion of polarized neon ions to the tip of the neon ion gas-field injection source can cause discontinuous changes in the beam current. Although the neon beam microscope could provide a number of functional advantages over the helium ion microscope [7], it has been

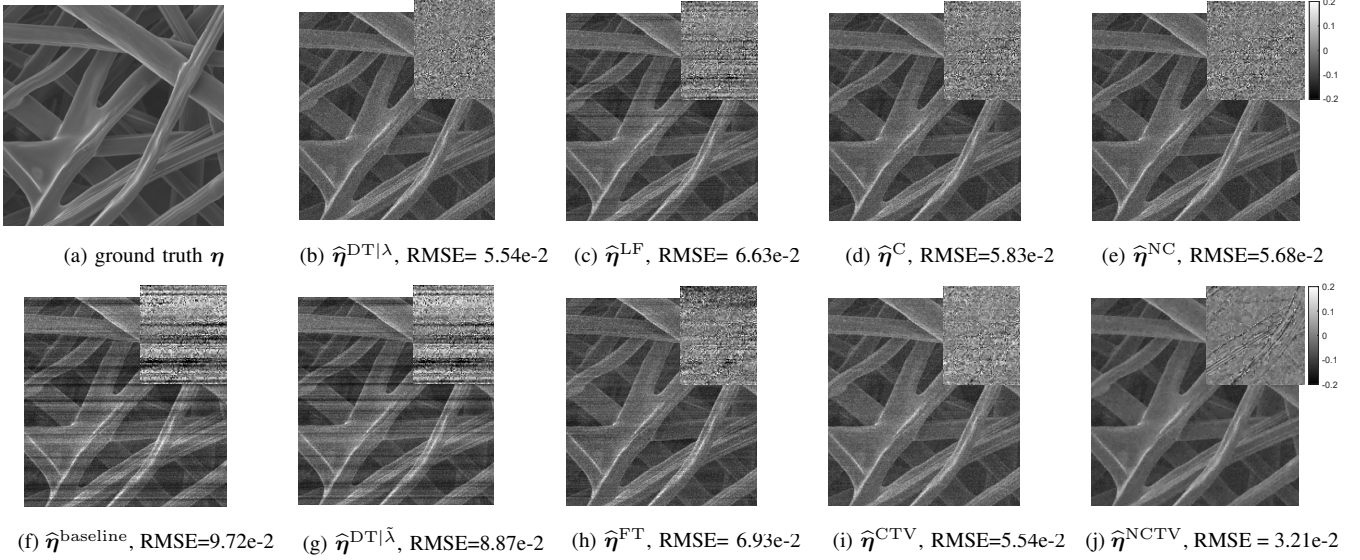


Fig. 7: SEM example with ground truth $\eta \in [0.2, 1]$, mean dose $\bar{\lambda} = 200$ and nominal sub-acquisition dose $\tilde{\lambda}/n = 0.1$. The actual dose λ is a Gaussian autoregressive process with correlation coefficient of 0.999 for neighboring pixels in a row and coefficient of variation $\sigma_\lambda/\bar{\lambda} = 0.2$. All micrograph images are on the same scale shown in (a), chosen so that no more than 2% of pixels are saturated in any given image. Inset images show error $\hat{\eta} - \eta$ for a subset of the image taken from the bottom middle of the image. Tuning parameters are: $\beta_{\text{NC}} = 2000$, $\beta_{\text{NCTV}} = 8$, $\beta_{\text{C}} = 100$, and $\beta_{\text{CTV}} = 3e - 4$. The non-causal estimator found $\hat{\lambda} = 202.75$ and the non-causal estimator with TV regularization found $\hat{\lambda} = 203.69$; the beam current empirical mean was $\frac{1}{p} \sum_{k=1}^p \lambda_k = 201.80$.

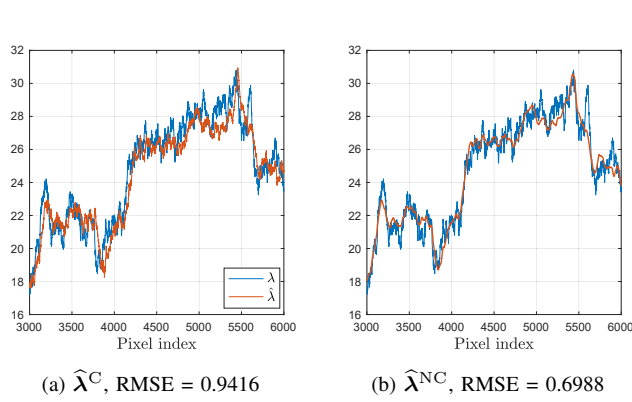


Fig. 8: Beam current estimates for a representative subset of pixels for the HIM example in Figure 6.

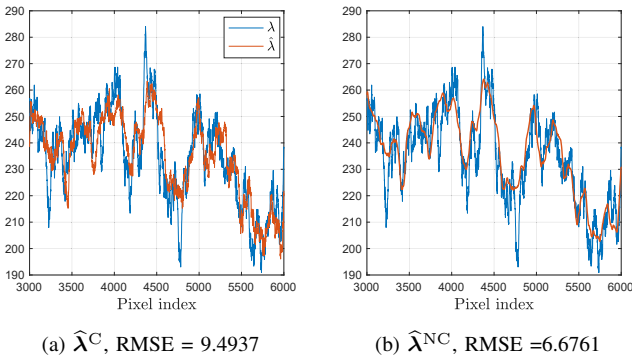


Fig. 9: Beam current estimates for a representative subset of pixels for the SEM example in Figure 7.

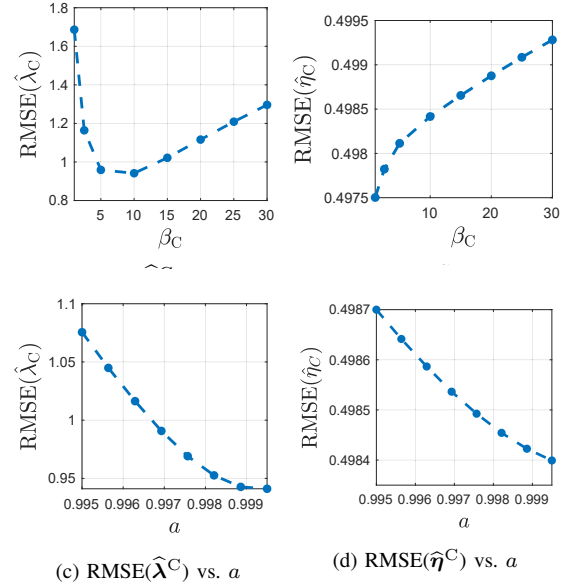


Fig. 10: Performance of causal joint estimation algorithm on data from Figure 6 for different values of a and β_{C} . For (a) and (b) correlation is fixed at $a = 0.999$; for (c) and (d), tuning parameter is fixed at $\beta_{\text{C}} = 10$.

less widely adopted because of difficulties maintaining a stable beam current.

A. A Discrete Markov Chain Model for Beam Current

The beam current in a neon beam microscope may be modeled using a two-state hidden Markov model (HMM). We assume a two-state model, and that the nature of the beam current variation has been well characterized so that the states

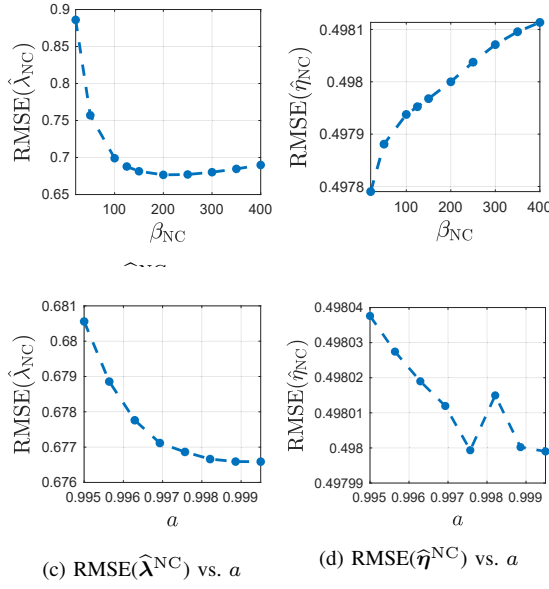


Fig. 11: Performance of non-causal joint estimation algorithm on data from Figure 6 for different values of a and β_{NC} . For (a) and (b), correlation is fixed at $a = 0.999$; for (c) and (d), tuning parameter is fixed $\beta_{NC} = 200$.

$\mathbf{s} \in \{s_1, s_2\}$ and transition probabilities $q(s, r) = P(\lambda_{k+1} = s | \lambda_k = r)$ are known. The mean beam current under this model is denoted $\bar{\lambda}$. Based on this model, we propose causal and non-causal joint estimation algorithms for η and λ .

B. Causal Estimation

Algorithm 1 describes our causal joint estimator $(\hat{\eta}^{HMM C}, \hat{\lambda}^{HMM C})$. At each pixel, $\hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \bar{\lambda})$ is used to form an initial estimate of η_k . As shown in [15] and [16], as well as in our own results in Section V-G, $\hat{\eta}_k^{DT|\bar{\lambda}}$ is actually quite close to the true η value, even when the beam current is imperfectly known. Thus, we initially assume that $\eta_k \approx \hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \bar{\lambda})$ and use the Forward algorithm [28], [29], as described in Algorithm 2, to compute the belief state $F_k(s)$ of λ_k given the measurements from that pixel and all previous pixels:

$$F_k(s) := P(\lambda_k = s | \mathbf{y}_{1:k}).$$

We pick $\hat{\lambda}_k^{HMM C}$ to be the state that maximizes $F_k(s)$. The estimate $\hat{\eta}_k^{HMM C}$ is produced by recomputing $\hat{\eta}_k^{DT|\bar{\lambda}}$, using $\bar{\lambda} = \hat{\lambda}_k^{HMM C}$ (Algorithm 1, Line 4). Note that Algorithm 2 operates recursively, requiring knowledge of $F_{k-1}(s)$ to compute $F_k(s)$. At the first pixel,

$$P(\mathbf{Y}_1 = \mathbf{y}_1 | \lambda_1 = s) = P_{\mathbf{Y}_k}(\mathbf{y}_1 | \lambda_k = s, \hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \bar{\lambda})),$$

where $P_{\mathbf{Y}_k}(\cdot; \cdot, \cdot)$ is the PMF in (8) and we have assumed that $\eta_k \approx \hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \bar{\lambda})$. It follows from the law of total probability that

$$P(\mathbf{Y}_1 = \mathbf{y}_1) = \sum_{s \in \mathbf{s}} P(\mathbf{Y}_1 = \mathbf{y}_1 | \lambda_1 = s) P(\lambda_1 = s),$$

where $P(\lambda_1 = s)$ is the stationary distribution of the hidden Markov chain. Applying Bayes's theorem, we find the initial belief state:

$$\begin{aligned} F_1(s) &= P(\lambda_1 = s | \mathbf{Y}_1 = \mathbf{y}_1) \\ &= \frac{P(\mathbf{Y}_1 = \mathbf{y}_1 | \lambda_1 = s) P(\lambda_1 = s)}{P(\mathbf{Y}_1 = \mathbf{y}_1)}. \end{aligned}$$

Algorithm 1 Causal joint estimation when beam current is modeled as a two-state hidden Markov chain

Input: $\mathbf{y}$, $\mathbf{s}$, $q(s, r) \forall s, r \in \mathbf{s}$, $F_1(s) \forall s \in \mathbf{s}$

- 1: **for** $k = [2, 3, \dots, p]$ **do**
 - 2: compute $F_k(s)$ using Algorithm 2
 - with $\eta_k = \hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \bar{\lambda})$ and $\mathbf{Y}_k = \mathbf{y}_k$
 - 3: $\hat{\lambda}_k^{HMM C} = \arg \max_{s \in \mathbf{s}} F_k(s)$
 - 4: $\hat{\eta}_k^{HMM C} = \hat{\eta}_k^{DT|\bar{\lambda}}(\bar{\lambda} = \hat{\lambda}_k^{HMM C})$
 - 5: **end for**
 - 6: **return** $\hat{\eta}^{HMM C}$, $\hat{\lambda}^{HMM C}$
-

C. Non-Causal Estimation

Our non-causal joint estimate $\hat{\lambda}_k^{HMM NC}$ selects the state with the greatest probability given the *entire* measurement sequence:

$$\hat{\lambda}_k^{HMM NC} = \arg \max_{s \in \mathbf{s}} P(\lambda_k = s | \mathbf{y}). \quad (47)$$

It uses the Forward-backward algorithm [28], [29] to compute a quantity proportional to $P(\lambda_k = s | \mathbf{y})$ for each of the two possible states. Note that $P(\lambda_k = s | \mathbf{y})$ may be factored as follows:

$$\begin{aligned} P(\lambda_k = s | \mathbf{y}) &\propto P(\lambda_k = s, \mathbf{y}) \\ &\stackrel{(a)}{=} P(\lambda_k = s, \mathbf{y}_{1:k}, \mathbf{y}_{k+1:p}) \\ &\stackrel{(b)}{=} P(\mathbf{y}_{k+1:p} | \lambda_k = s, \mathbf{y}_{1:k}) P(\lambda_k = s, \mathbf{y}_{1:k}) \\ &\stackrel{(c)}{=} \underbrace{P(\mathbf{y}_{k+1:p} | \lambda_k = s)}_{\text{backward pass}} \underbrace{P(\lambda_k = s, \mathbf{y}_{1:k})}_{\text{forward pass}}, \end{aligned} \quad (48)$$

where (a) follows from splitting the components of $\mathbf{y}$; (b) from the multiplication rule; and (c) from the conditional independence of $P(\mathbf{y}_{k+1:p} | \lambda_k = s)$ from $\mathbf{y}_{1:k}$. The final two factors include a term readily available from the *forward* pass of our causal algorithm: $P(\lambda_k = s, \mathbf{y}_{1:k}) \propto F_k(s)$, and a second term computed in a new recursive *backward* pass over the data. This new term,

$$B_k(s) := P(\mathbf{y}_{k+1:p} | \lambda_k = s),$$

may be computed using the following recursive formula moving backwards over the data sequence:

$$B_k(s) = \sum_{s' \in \mathbf{s}} B_{k+1}(s') P_{\mathbf{Y}_k}(\mathbf{y}_{k+1} | \lambda_{k+1} = s') q(s, s').$$

As stipulated by the Forward-backward algorithm, the last pixel is initialized with $B_n(s) = 1 \forall s \in \mathbf{s}$. Just as in the causal algorithm described in Algorithm 1, our non-causal joint estimation algorithm forms an initial estimate of η_k at each

Algorithm 2 Forward Algorithm for computing belief state $F_k(s)$ at k th pixel given η_k

Input: $\mathbf{Y}_k = \mathbf{y}_k \in \mathbb{R}^n$, η_k , $\mathbf{s}$, $q(s, r) \forall s, r \in \mathbf{s}$, $F_{k-1}(s) \forall s \in \mathbf{s}$

- 1: Compute $P(\lambda_k = s | \text{all past measurements}) = \sum_{s' \in \mathbf{s}} q(s', s) F_{k-1}(s') \forall s \in \mathbf{s}$
 - 2: Compute $P_{\mathbf{Y}_k}(\mathbf{y}_k | \eta_k, \lambda_k = s) \forall s \in \mathbf{s}$
 - 3: $P(\mathbf{Y}_k = \mathbf{y}_k | \text{all past measurements}) = \sum_{s \in \mathbf{s}} P_{\mathbf{Y}_k}(\mathbf{y}_k | \eta_k, \lambda_k = s) P(\lambda_k = s | \text{all past measurements})$
 - 4: $F_k(s) = P_{\mathbf{Y}_k}(\mathbf{y}_k | \eta_k, \lambda_k = s) P(\lambda_k = s | \text{all past measurements}) / P(\mathbf{Y}_k = \mathbf{y}_k | \text{all past measurements})$
 - 5: **return** $F_k(s)$
-

pixel using $\hat{\eta}_k^{\text{DT}|\tilde{\lambda}}(\tilde{\lambda} = \bar{\lambda})$. Assuming $\eta_k \approx \hat{\eta}_k^{\text{DT}|\tilde{\lambda}}(\tilde{\lambda} = \bar{\lambda})$, the estimate $\hat{\lambda}_k^{\text{HMM NC}}$ is formed according to (47), requiring a forward and backward pass to compute the two terms in (48). Then, η is estimated according to: $\hat{\eta}_k^{\text{HMM NC}} = \hat{\eta}_k^{\text{DT}|\tilde{\lambda}}(\tilde{\lambda} = \hat{\lambda}_k^{\text{HMM NC}})$.

D. Simulated Microscopy Results

Synthetic measurements were generated using an existing micrograph as the ground truth image. The beam current time series was produced according to a two-state Markov chain model with $\lambda \in \{20, 30\}$ using transition probabilities

$$\begin{aligned} P(\lambda_k = 20 | \lambda_{k-1} = 30) &= 0.003, & \text{and} \\ P(\lambda_k = 30 | \lambda_{k-1} = 20) &= 0.002, \end{aligned}$$

resulting in $\bar{\lambda} = 24$. At each pixel, the dwell time was divided into $n = 300$ sub-acquisitions.

In Figure 12, we compare the RMSE results for the following methods:

- baseline: $\hat{\eta}^{\text{baseline}}$ is the pixel-wise evaluation of (3) independently at each pixel using assumed dose $\bar{\lambda}$.
- DT| λ : $\hat{\eta}^{\text{DT}|\lambda}$ is the pixel-wise ML estimate (9) computed with true beam current λ (provided by an oracle).
- DT| $\tilde{\lambda}$: $\hat{\eta}^{\text{DT}|\tilde{\lambda}}$ is the pixel-wise ML estimate (9) computed using assumed dose $\tilde{\lambda} = \bar{\lambda}$.
- HMM causal: $(\hat{\eta}^{\text{HMM C}}, \hat{\lambda}^{\text{HMM C}})$, computed using Algorithm 2.
- HMM non-causal: $(\hat{\eta}^{\text{HMM NC}}, \hat{\lambda}^{\text{HMM NC}})$, computed according to Section VI-C.

The baseline estimate $\hat{\eta}^{\text{baseline}}$ in Figure 12d exhibits prominent stripe artifacts. Figure 12e shows $\hat{\eta}^{\text{DT}|\tilde{\lambda}}$ with stripe artifacts greatly reduced but still visible. Figures 12c and 12f show results for our causal and non-causal HMM joint estimation algorithms. In both cases, RMSE is further reduced over $\hat{\eta}^{\text{DT}|\tilde{\lambda}}$, approaching the performance of $\hat{\eta}^{\text{DT}|\lambda}$ in Figure 12b, with $\hat{\eta}^{\text{HMM NC}}$ slightly outperforming $\hat{\eta}^{\text{HMM C}}$.

In Figure 12g, we plot the true beam current time series λ with estimates $\hat{\lambda}^{\text{HMM C}}$ and $\hat{\lambda}^{\text{HMM NC}}$. Note that both estimates match the true beam current at the vast majority of pixels, with an error percentage of 0.89% for the causal algorithm and 0.21% for the non-causal algorithm. Although performance is very good with only the causal forward pass, approximately four times fewer errors occur in the λ estimate when all data is considered.

VII. CONCLUSION

In this work, we explore the estimation of *two* properties at each pixel of a particle beam micrograph: mean SE yield

η and beam current λ . Using the Cramér–Rao bound at a single pixel, we show the feasibility of joint estimation given time-resolved measurements. Specifically, we show that at high η , joint estimation is only slightly more challenging than estimating η when λ is given. We demonstrate that when the dose is sufficiently high, joint estimation is possible at even a single pixel. To perform joint estimation at moderate doses, we exploit the fact that beam current does not vary arbitrarily. The algorithms of Section V are motivated by electron and helium ion beams, where current is smoothly varying. Algorithms in Section VI are designed for neon beam microscopes, where beam current is known to jump among known values. Through tests performed on synthetic microscopy data, we show that our η estimators outperform existing methods and our novel λ estimators closely match the ground truth.

Our innovation not only prevents micrograph artifacts that arise when the beam current is not perfectly known, but also provides the operator with new and useful information. Knowledge of the beam current could save costly instrument maintenance time, enable on-the-fly instrument control, improve micrographs and milling outcomes, and even further the proliferation of powerful new instruments like the neon beam microscope, where maintaining a stable beam current is a key challenge. The promise of time-resolved sensing extends beyond online beam current estimation. Although we are exploiting measurements at the time scale of subacquisitions, the estimates in this paper are all formed at the (coarser) per-pixel level. Estimates at the (finer) subacquisition level may enable partial compensations for drift, sample charging, and contamination.

APPENDIX

In gradient algorithms, we require methods to evaluate or approximate derivatives of the logarithm of the PMF $P_Y(y; \eta, \lambda)$ in (1). The $y = 0$ case is mathematically simple and also important because many sub-acquisitions result in no detected SEs. For $y > 0$, elegant expressions can be given using Touchard polynomials. With discrete-time data with sufficiently large n , the relevant sub-acquisition dose is small, so we also derive approximations that hold for small λ .

$y = 0$ Case

By substitution and simplification,

$$\log P_Y(0; \eta, \lambda) = -\lambda(1 - e^{-\eta}). \quad (49)$$

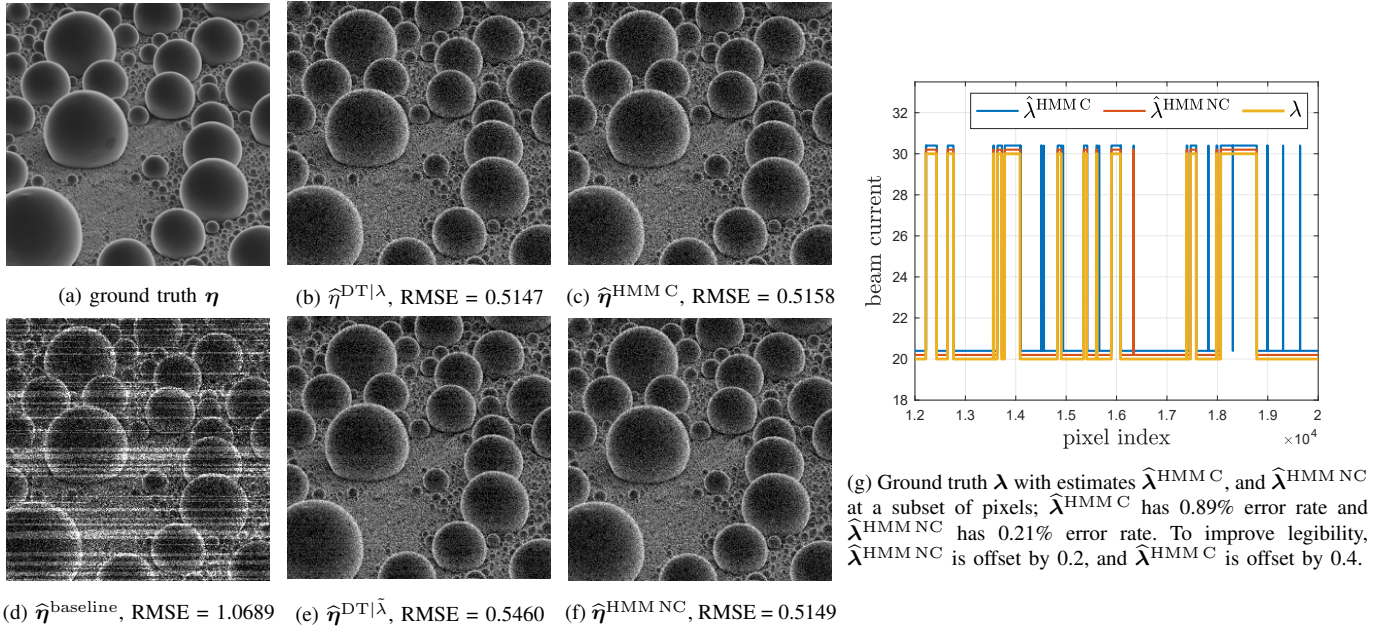


Fig. 12: Results from a synthetic neon beam experiment with $\eta \in [2, 6]$, $n = 300$, and λ modeled as a two-state hidden Markov chain with $\lambda \in \{20, 30\}$.

The derivatives of this for optimization over λ or η are

$$\frac{d}{d\lambda} [-\lambda(1 - e^{-\eta})] = -(1 - e^{-\eta}), \quad (50a)$$

$$\frac{d^2}{d\lambda^2} [-\lambda(1 - e^{-\eta})] = 0, \quad (50b)$$

$$\frac{d}{d\eta} [-\lambda(1 - e^{-\eta})] = -\lambda e^{-\eta}, \quad (50c)$$

$$\frac{d^2}{d\eta^2} [-\lambda(1 - e^{-\eta})] = \lambda e^{-\eta}. \quad (50d)$$

Touchard Polynomials

The *Touchard polynomials* are defined by

$$T_n(x) = \sum_{k=0}^n S(n, k) x^k, \quad (51)$$

where $S(n, k)$ is a *Stirling number of the second kind*, i.e., the number of partitions of a set of size n into k disjoint non-empty subsets. Stirling numbers of the second kind can be used to write

$$m^y = \sum_{k=0}^y S(y, k) \frac{m!}{(m-k)!},$$

where we regard $1/(m-k)! = 0$ if $k > m$ [30]. Then

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{m^y}{m!} x^m &= \sum_{m=0}^{\infty} \left(\sum_{k=0}^y S(y, k) \frac{m!}{(m-k)!} \right) \frac{x^m}{m!} \\ &= \sum_{k=0}^y S(y, k) \sum_{m=0}^{\infty} \frac{x^m}{(m-k)!} = \sum_{k=0}^y S(y, k) x^k e^x \\ &= T_y(x) e^x. \end{aligned} \quad (52)$$

Now we have

$$\frac{d}{dx} \sum_{m=0}^{\infty} \frac{m^y}{m!} x^m \stackrel{(a)}{=} \sum_{m=0}^{\infty} \frac{m^{y+1}}{m!} x^{m-1} \stackrel{(b)}{=} \frac{1}{x} T_{y+1}(x) e^x, \quad (53)$$

where (a) follows from term-by-term differentiation; and (b) from (52). It is the derivative of the log that will be useful in what follows:

$$\frac{d}{dx} \log \sum_{m=0}^{\infty} \frac{m^y}{m!} x^m = \frac{T_{y+1}(x)}{x T_y(x)}, \quad (54)$$

which now follows from the chain rule and substitution of (52) and (53).

From (51), a good approximation to $T_n(x)$ for small x can be obtained by truncating to $k \in \{0, 1, 2\}$. While Stirling numbers of the second kind $S(y, k)$ are not easy to work with in general, for any $y \geq 1$, we have $S(y, 0) = 0$, $S(y, 1) = 1$, and $S(y, 2) = 2^{y-1} - 1$. Therefore, for any $y \geq 1$,

$$T_y(x) \approx x + (2^{y-1} - 1)x^2, \quad (55a)$$

$$T'_y(x) \approx 1 + (2^y - 2)x, \quad (55b)$$

for small x .

Derivatives of log likelihood with respect to λ

We have

$$\begin{aligned} \frac{d}{d\lambda} \log \left[\frac{e^{-\lambda\eta^y}}{y!} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &= -1 + \frac{d}{d\lambda} \log \left[\sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &\stackrel{(a)}{=} -1 + e^{-\eta} \frac{T_{y+1}(\lambda e^{-\eta})}{\lambda e^{-\eta} T_y(\lambda e^{-\eta})} = -1 + \frac{T_{y+1}(\lambda e^{-\eta})}{\lambda T_y(\lambda e^{-\eta})}, \end{aligned} \quad (56)$$

where (a) follows from the chain rule and (54). It follows that

$$\begin{aligned} & \frac{d^2}{d\lambda^2} \log \left[\frac{e^{-\lambda} \eta^y}{y!} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &= \frac{e^{-\eta} T'_{y+1}(\lambda e^{-\eta})}{\lambda T_y(\lambda e^{-\eta})} - \frac{T_{y+1}(\lambda e^{-\eta})}{\lambda^2 T_y(\lambda e^{-\eta})} \\ & \quad - \frac{\lambda e^{-\eta} T_{y+1}(\lambda e^{-\eta}) T'_y(\lambda e^{-\eta})}{(\lambda T_y(\lambda e^{-\eta}))^2}. \end{aligned} \quad (57)$$

Substituting the second-order approximation (55) in (56) and (57) gives the approximations

$$\frac{d}{d\lambda} [\sim] \approx -1 + \frac{1}{\lambda} \cdot \frac{1 + (2^y - 1)\lambda e^{-\eta}}{1 + (2^{y-1} - 1)\lambda e^{-\eta}}, \quad (58a)$$

$$\frac{d^2}{d\lambda^2} [\sim] \approx -\frac{1}{\lambda^2} \left[1 + \frac{(2^y - 1)2^{y-1}(\lambda e^{-\eta})^2}{(1 + (2^{y-1} - 1)\lambda e^{-\eta})^2} \right]. \quad (58b)$$

Derivatives of log likelihood with respect to η

We have

$$\begin{aligned} & \frac{d}{d\eta} \log \left[\frac{e^{-\lambda} \eta^y}{y!} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &= \frac{y}{\eta} + \frac{d}{d\eta} \log \left[\sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &\stackrel{(a)}{=} \frac{y}{\eta} - \lambda e^{-\eta} \frac{T_{y+1}(\lambda e^{-\eta})}{\lambda e^{-\eta} T_y(\lambda e^{-\eta})} = \frac{y}{\eta} - \frac{T_{y+1}(\lambda e^{-\eta})}{T_y(\lambda e^{-\eta})}, \end{aligned} \quad (59)$$

where (a) follows from the chain rule and (54). It follows that

$$\begin{aligned} & \frac{d^2}{d\eta^2} \log \left[\frac{e^{-\lambda} \eta^y}{y!} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m m^y}{m!} \right] \\ &= -\frac{y}{\eta^2} + \lambda e^{-\eta} \frac{T'_{y+1}(\lambda e^{-\eta})}{T_y(\lambda e^{-\eta})} \\ & \quad - \lambda e^{-\eta} \frac{T_{y+1}(\lambda e^{-\eta}) T'_y(\lambda e^{-\eta})}{(T_y(\lambda e^{-\eta}))^2} \\ &= -\frac{y}{\eta^2} + \lambda e^{-\eta} \\ & \quad \times \frac{T'_{y+1}(\lambda e^{-\eta}) T_y(\lambda e^{-\eta}) - T_{y+1}(\lambda e^{-\eta}) T'_y(\lambda e^{-\eta})}{(T_y(\lambda e^{-\eta}))^2}. \end{aligned} \quad (60)$$

Substituting the second-order approximation (55) in (59) and (60) gives the approximations

$$\frac{d}{d\eta} [\sim] \approx \frac{y}{\eta} - \frac{1 + (2^y - 1)\lambda e^{-\eta}}{1 + (2^{y-1} - 1)\lambda e^{-\eta}}, \quad (61a)$$

$$\frac{d^2}{d\eta^2} [\sim] \approx -\frac{y}{\eta^2} + \frac{2^{y-1} \lambda e^{-\eta}}{(1 + (2^{y-1} - 1)\lambda e^{-\eta})^2}. \quad (61b)$$

ACKNOWLEDGMENT

The authors thank Sangchul Lee and Claude Leibovici for answers to a Stack Exchange query [30], [31] that were instrumental to the derivations in the appendix.

REFERENCES

- [1] B. W. Ward, J. A. Notte, and N. P. Economou, "Helium ion microscope: A new tool for nanoscale microscopy and metrology," *J. Vacuum Sci. Technol. B*, vol. 24, no. 6, pp. 2871–2874, Nov. 2006.
- [2] J. Notte and J. Huang, *The Helium Ion Microscope*. Cham: Springer International Publishing, 2016, pp. 3–30.
- [3] J. Morgan, J. Notte, R. Hill, and B. Ward, "An introduction to the helium ion microscope," *Microscopy Today*, vol. 14, no. 4, pp. 24–31, 2006.
- [4] F. F. Rahman, J. A. Notte, R. H. Livengood, and S. Tan, "Observation of synchronized atomic motions in the field ion microscope," *Ultramicroscopy*, vol. 126, pp. 10–18, 2013.
- [5] A. J. Barlow, J. F. Portoles, N. Sano, and P. J. Cumpson, "Removing beam current artifacts in helium ion microscopy: A comparison of image processing techniques," *Microsc. Microanal.*, vol. 22, no. 5, pp. 939–947, Oct. 2016.
- [6] J. Notte, F. Rahman, S. McVey, S. Tan, and R. Livengood, "The neon gas field ion source - stability and lifetime," *Microsc. Microanal.*, vol. 16, no. S2, p. 28–29, 2010.
- [7] F. H. M. Rahman, S. McVey, L. Farkas, J. A. Notte, S. Tan, and R. H. Livengood, "The prospects of a subnanometer focused neon ion beam," *Scanning*, vol. 34, no. 2, pp. 129–134, 2012.
- [8] A. J. Barlow, N. Sano, B. J. Murdoch, J. F. Portoles, P. J. Pigram, and P. J. Cumpson, "Observing the evolution of regular nanostructured indium phosphide after gas cluster ion beam etching," *Applied Surface Science*, vol. 459, pp. 678–685, 2018.
- [9] J. Schwartz, Y. Jiang, Y. Wang, A. Aiello, P. Bhattacharya, H. Yuan, Z. Mi, N. Bassim, and R. Hovden, "Removing stripes, scratches, and curtaining with nonrecoverable compressed sensing," *Microsc. Microanal.*, vol. 25, no. 3, p. 705–710, 2019.
- [10] A. Khalilian-Gourtani, M. Tepper, V. Minden, and D. B. Chklovskii, "Strip the stripes: Artifact detection and removal for scanning electron microscopy imaging," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2019, pp. 1060–1064.
- [11] S. Liu, L. Sun, J. Gao, and K. Li, "A fast curtain-removal method for 3D FIB-SEM images of heterogeneous minerals," *J. Microscopy*, vol. 272, no. 1, pp. 3–11, 2018.
- [12] T. Ott, D. Roldán, C. Redenbach, K. Schladitz, M. Godehardt, and S. Höhn, "Three-dimensional structural comparison of tantalum glancing angle deposition thin films by FIB-SEM," *J. Sensors Sensor Syst.*, vol. 8, no. 2, pp. 305–315, 2019.
- [13] M. Peng, J. Murray-Bruce, K. K. Berggren, and V. K. Goyal, "Source shot noise mitigation in focused ion beam microscopy by time-resolved measurement," *Ultramicroscopy*, vol. 211, p. 112948, 2020.
- [14] M. Peng, J. Murray-Bruce, and V. K. Goyal, "Time-resolved focused ion beam microscopy: Modeling, estimation methods, and analyses," *IEEE Trans. Computational Imaging*, vol. 7, pp. 547–561, 2021.
- [15] L. Watkins, S. W. Seidel, M. Peng, A. Agarwal, C. C. Yu, and V. K. Goyal, "Robustness of time-resolved measurement to unknown and variable beam current in particle beam microscopy," in *Proc. IEEE Int. Conf. Image Process.*, Anchorage, AK, Sep. 2021, pp. 3487–3491.
- [16] —, "Prevention beats removal: Avoiding stripe artifacts from current variation in particle beam microscopy through time-resolved sensing," *Microsc. Microanal.*, vol. 27, no. S1, pp. 422–425, Aug. 2021.
- [17] P. Cizmar, A. E. Vladár, and M. T. Postek, "Real-time scanning charged-particle microscope image composition with correction of drift," *Microsc. Microanal.*, vol. 17, no. 2, pp. 302–308, Apr. 2011.
- [18] S. Yamada, T. Ito, K. Gouhara, and Y. Uchikawa, "Electron-count imaging in SEM," *Scanning*, vol. 13, no. 2, pp. 165–171, 1991.
- [19] L. Jin, A.-C. Milazzo, S. Kleinfelder, S. Li, P. Leblanc, F. Duttweiler, J. C. Bouwer, S. T. Peltier, M. H. Ellisman, and N.-H. Xuong, "Applications of direct detection device in transmission electron microscopy," *J. Structural Biology*, vol. 161, no. 3, pp. 352–358, 2008.
- [20] G. McMullan, S. Chen, R. Henderson, and A. R. Faruqi, "Detective quantum efficiency of electron area detectors in electron microscopy," *Ultramicroscopy*, vol. 109, no. 9, pp. 1126–1143, 2009.
- [21] A. Agarwal, J. Simonaitis, V. K. Goyal, and K. K. Berggren, "Secondary electron count imaging in SEM," arXiv:2111.01862, Nov. 2021.
- [22] K. S. Sim, J. T. L. Thong, and J. C. H. Phang, "Effect of shot noise and secondary emission noise in scanning electron microscope images," *Scanning*, vol. 26, no. 1, pp. 36–40, 2004.
- [23] A. Agarwal, J. Simonaitis, and K. K. Berggren, "Image-histogram-based secondary electron counting to evaluate detective quantum efficiency in SEM," *Ultramicroscopy*, vol. 224, p. 113238, 2021.
- [24] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

- [25] J. Notte, B. Ward, N. Economou, R. Hill, R. Percival, L. Farkas, and S. McVey, “An introduction to the helium ion microscope,” in *AIP Conference Proceedings*, vol. 931. AIP, 2007, pp. 489–496.
- [26] Y. Lin and D. C. Joy, “A new examination of secondary electron yield data,” *Surface Interface Anal.*, vol. 37, no. 11, pp. 895–900, 2005.
- [27] L. Watkins, “Mitigating current variation in particle beam microscopy,” arXiv:2106.04686, Jun. 2021, B.S. honors thesis in electrical and computer engineering, Boston University.
- [28] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” *Annals Math. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.
- [29] L. E. Baum, “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes,” *Inequalities*, vol. 3, no. 1, pp. 1–8, 1972.
- [30] S. Lee, Mathematics Stack Exchange, Jul. 2020, <https://math.stackexchange.com/q/3744576>.
- [31] C. Leibovici, Mathematics Stack Exchange, Jul. 2020, <https://math.stackexchange.com/q/3744557>.