Abstract—Photon counting lidar has become an invaluable tool for 3D depth imaging due to the fine depth precision it can achieve over long ranges, with emerging applications in robotics, autonomous vehicles and remote sensing. However, high frame rate, high resolution lidar devices produce an enormous amount of time-of-flight (ToF) data which can cause a severe data processing bottleneck hindering the deployment of real-time systems. In this paper, we show that this bottleneck can be avoided through the use of a hardware-friendly compressed statistic, or a so-called spline sketch, of the ToF data, massively reducing the data rate without sacrificing the quality of the recovered depth image. Specifically, we present previously proposed Fourier sketches, piecewise linear or quadratic spline sketches are able to reconstruct real-world depth images with negligible loss of resolution whilst achieving 95% compression compared to the full ToF data, as well as offering multi-peak detection performance. However, unlike Fourier sketches, spline sketches require minimal on-chip arithmetic computation per photon detection. We also show that by building in appropriate range-walk correction, spline sketches can be made robust to photon pile-up effects associated with bright reflectors. We contrast this with previously proposed solutions such as coarse binning histograms that trade depth resolution for data compression, suffer from a highly nonuniform accuracy across depth and can fail catastrophically when imaging bright reflectors. By providing a practical means of overcoming the data processing bottleneck, spline sketches offer a promising route to low cost high rate, high resolution lidar imaging.

Index Terms—Single-photon Lidar, data compression, compressive Learning, splines, Cramér-Rao bounds.

I. INTRODUCTION

SINGLE-PHOTON light detection and ranging (lidar) is a well-established 3D depth imaging modality that offers millimeter precision [1] over long ranges [2]. The technique consists of emitting light pulses and using a single photon avalanche diode (SPAD) to detect the presence of an incoming photon. By using a time-correlated single photon counting system (TCSPC), time-of-flight data can be accumulated over a series of light pulses providing depth and intensity information for the surfaces in each pixel of the scene. A TCSPC histogram is most commonly used to represent the ToF data by clustering the time delays between emitted light pulses and detected photons into time bins discretized over the whole timing depth window for each pixel in the scene (see Fig. 1). Rapid hardware and technological advances in recent years have granted even finer resolution at much higher frame rates [3], [4]. Thus modern lidar devices generate massive amounts of data per second that needs to be transferred off-chip for downstream tasks such as detection, depth estimation and segmentation. As an example, a high rate, high resolution lidar device is capable of imaging a scene containing $512 \times 512$ pixels at a frame rate of 50 frames per second (fps). Assuming the timing depth window of the laser is discretized over $T = 1000$ histogram bins of 16 b precision, the lidar device would require a data transfer rate of almost 15 GB/s.

To tackle the data-transfer bottleneck of modern lidar devices, practitioners have often resorted to using coarse timing bin width in the TCSPC histogram. This technique is often referred to as coarse binning and typically results in a substantial loss of temporal resolution causing a limiting compression-resolution trade-off as demonstrated in Fig. 1. Gyongy et al. [5] proposed a practical solution by increasing the width of the impulse response of the device and using a maximum likelihood estimator to achieve sub-bin resolution with respect to coarse bins. However, as will be demonstrated in Section VI, we observe that 1) the worst-case error is typically highly dependent on the position of the signal with respect to the coarse bins, and 2) if there are highly reflective surfaces (e.g. retroreflectors) in the scene, the measured IRF can be much narrower than the system’s normal impulse response therefore sub-bin accuracy with respect to the
coarse bins cannot be achieved. Recently, Sheehan et al. [6], [7], [8] proposed a novel solution to tackle the data-transfer bottleneck whilst circumventing the compression-resolution trade-off. The technique consists of forming a compact representation, a so-called sketch, of the ToF data that retains sufficient salient information to accurately estimate the depth and intensity of multiple surfaces in the scene. Fundamentally, it was demonstrated that the size of the sketch only needs to be of the order of the number of surfaces in the scene to achieve negligible loss of resolution [6]. As there are typically no more than 1 or 2 surfaces present per pixel, a substantial compression rate can be achieved. Furthermore, the reconstruction algorithms are based solely on the sketch and admit both a computational and memory complexity that scales only on the size of the sketch and is independent of both the resolution of the lidar device as well as the number of photons detected. However, the specific sketch proposed in [6] requires the on-chip computation of several sinusoidal functions for each photon detection which may be challenging to execute within the limited dead-time (a period of insensitivity after a photon detection during which the SPAD cannot register any incoming photons) of the lidar device.

In this paper, we introduce new sketch statistics that both capture the necessary salient information in the data while requiring low on-chip arithmetic complexity. To this end, we propose a spline sketch approach for photon counting that unites the substantial data transfer compression capabilities of the Fourier sketched lidar technique with the hardware-friendly computation of a traditional TCSPC histogram. By replacing the sinusoidal functions of the original sketch in [6] by elementary piecewise polynomial spline functions, we can construct a compressed representation of the ToF data that is almost as statistically efficient as the originally proposed Fourier sketch while having minimal computational overhead on-chip. The work presented here builds upon our initial conference paper [9] where the spline sketch was first proposed. In the current work, we further calculate the on-chip computational cost for different sketching solutions in comparison with the traditional histogram, and derive algorithm-independent Cramér-Rao Bounds (CRB) for the performance of each approach, highlighting the strong spatial dependence of these for current histogram solutions. This theory is supported with experimental validation, showing that the proposed reconstruction algorithms perform roughly in line with the CRB predictions, as well as providing good empirical performance on real data. The main contributions of this paper are:

- Through the introduction of piecewise polynomial splines, we design low-cost, hardware-efficient sketches that retain sufficient salient information required to estimate the depth and intensity parameters of the lidar observation model and bridge the gap between coarse binning representations and Fourier sketches whose performance is position independent by definition.
- Both closed-form (for linear splines) and iterative algorithms are proposed that can accurately estimate the depth and intensity parameters of the scene while exhibiting a computational and memory complexity that scales solely with the size of the spline sketch.
- We show that the performance of coarse binning is highly dependent on the position of the reflector within a given bin. In contrast, linear and quadratic polynomial spline sketches exhibit only mild variation in such performance and achieve theoretical error almost equivalent to the Fourier sketch.
- Finally, we show that the spline sketches carry sufficient information to enable the correction of range walk associated with bright reflective surfaces and present a prototype spline sketch range walk correction algorithm.

The paper is organized as follows: In Section II the concepts of sketched lidar and piecewise polynomial splines are introduced. In Section III, the cost of computing the spline sketches on-chip is discussed while the spline sketch-based reconstruction algorithms are proposed in Section IV. Both theoretical Cramér-Rao error bounds and the relative performance of the proposed algorithms are demonstrated in Sections V and VI, respectively. Finally, we conclude the paper in Section VII.

II. BACKGROUND

A. Lidar Observation Model

During the course of $N$ laser pulses, the SPAD registers $n \leq N$ photon detections for a given pixel in the scene, where each photon has a detection time denoted by $x_j$ for $1 \leq j \leq n$. Assuming there are $K$ distinct surfaces in the field of view each at a timing depth of $t_k \in [0, T - 1]$, then a simple model for the detection time $x_j$ of a single photon for a given pixel is given by a mixture distribution $[10]^1$

$$
\pi(x_j|\theta) = \sum_{k=1}^{K} \alpha_k \pi_k(x_j|t_k) + \alpha_0 \pi_0(x_j),
$$

where $\alpha_1, \alpha_2, \ldots, \alpha_K$ denote the probability of detection from the $k$th surface, $\alpha_0$ denotes the probability of detection from an ambient background source (e.g. sunlight) with $\sum_{k=0}^{K} \alpha_k = 1$, and $\theta = \{\alpha_1, \alpha_2, \ldots, \alpha_K, t_1, \ldots, t_K\}$ denotes the set of 2 $K$ unknown parameters. We can then define the detection Signal-to-Background Ratio (SBR) as $\text{SBR} = (\sum_{k=1}^{K} \alpha_k)/\alpha_0$ (note that this is not the same as the raw sensor SBR which can be much lower $[11]$).

Given that the lidar system has a known impulse response function (IRF) $h(t)$ then the distribution of photons originating from a surface is defined as $\pi_s(x|t) = h(x - t)/H$ where $H = \sum_{t=0}^{T-1} h(t)$. The distribution of background photon detections can be modelled uniformly over the whole timing depth window $\pi_b(x) = 1/T$. As a result, single photon counting lidar reduces to estimating the parameters $\theta$ given a collection of ToF data (e.g. TCSPC histogram) for each pixel in the scene. As discussed in Section I, the complexities of transferring ToF data are often dependent on the depth resolution of the lidar device which is determined by the discrete number of time bins $T$ used over the whole timing depth window. The memory requirement for a

---

1This model does not capture all aspects of the SPAD device. However it serves as a useful idealisation. The sketching framework is not reliant on this specific model and indeed, a more complicated one will be considered in Section IV-C when discussing range walk effects.
single pixel in a frame is typically $\mathcal{O}(Tb)$ where $b$ denotes the bit precision per bin. This can be memory expensive in modern day devices and provides a barrier to developing high resolution systems.

B. Sketched Lidar

Recently, Sheehan et al. [6, 8] developed a novel solution to capture sufficient salient information from the ToF data $\{x_j\}_{j=1}^n$ to estimate the parameters of the observation model in (1). The TCSPC histogram (see Fig. 1) can be seen as a statistic that captures enough information from the ToF data required for estimation. However it is a very inefficient representation with memory size scaling proportional to required resolution. The authors in [6] designed a much more compact representation of the ToF data that has a size independent of the resolution $T$ of the lidar device and the number of photon detections $n$.

The basic concept of sketching is to introduce a nonlinear mapping $\Phi(x)$ that defines a parameterised sketch statistic, $z_0 = \mathbb{E}_{x \sim \pi(\cdot)} \Phi(x)$, and its empirical estimate, $\hat{z} = \frac{1}{n} \sum_{j=1}^n \Phi(x_j)$, with respect to data $x_j$ drawn from the probability density $\pi(x)$. Then we can estimate the parameters $\theta$ by comparing the empirical sketch, $\hat{z}$ with the hypothesised expectation: $\mathbb{E}_{x \sim \pi(\cdot)} \Phi(x)$. In the original sketched lidar work [6] a particular Fourier sketch was proposed that was inspired by compressive learning theory [12]. Let $\Phi_F(x) = |e^{i\omega x}|_{\ell=1}^m$ denote $m$ complex Fourier features, then the (complex) Fourier sketch of size $m$ can be defined as

$$\hat{z} = \frac{1}{n} \sum_{j=1}^n \Phi_F(x_j).$$

Due to the summation in (2), the sketch can be updated throughout the pulse cycle for each photon detection. Once the sketch has been computed, the parameters of the observation model in (1) can be estimated, e.g. through a generalized method of moments [13] scheme

$$\arg\min_\theta \|\hat{z} - \mathbb{E}_{x \sim \pi(\cdot)} \Phi_F(x)\|_W^2,$$

for some positive definite weighting matrix $W \in \mathbb{C}^{m \times m}$. Advantageously for the Fourier sketch, the sketch in expectation is simply the characteristic function (CF) of the observation model in (1) sampled at the $m$ frequencies $\omega_1, \ldots, \omega_m$. The CF exists for all distributions and, in the case of the observation model in (1), has a simple closed-form solution, which can be easily computed. Fundamentally, the size of the sketch (i.e. the number of CF samples) required to accurately estimate $\theta$ needs only to be of the order of the number of surfaces in the pixel, for instance $m = \mathcal{O}(K)$. Moreover, in [6], [8], the authors propose several sketch-only reconstruction algorithms, including exploiting powerful spatial denoisers, that have both a memory and computational complexity of $\mathcal{O}(m)$.

One of the challenges with the original sketched lidar approach is that the feature function $\Phi_F$ requires nontrivial on-chip processing. Firstly, the feature function in (2) consists of 2$m$ sinusoidal functions$^2$ which is resource intensive, despite the existence of various FPGA implementations, e.g. [14]. Secondly, the Fourier sketch is a global sketch as each sinusoidal function has a support equal to the whole pulse cycle and therefore all 2$m$ components of the sketch must be updated for each photon detection. This is in contrast to the TCSPC histogram where only a single timing bin needs to be updated for each photon detection. In this paper we propose to replace the Fourier features $\Phi_F$ with a low-cost, semi-local spline sketch consisting of $M$ piecewise polynomial spline feature functions that have a limited support over the timing depth window. While one can intuitively think of the spline functions as approximating the Fourier representation, it is important to stress that they define valid sketches in their own right (we are not simply approximating the Fourier sketch) and as we will see in Section IV-A have additional pleasing properties that are unrelated to the previous Fourier sketch.

C. Cardinal Basis Splines

Cardinal splines are piecewise polynomial functions of degree $p$ that are used extensively in approximation theory [15]. Let $\phi_0$ denote the spline of degree $p = 0$ defined as

$$\phi_0(x) = \begin{cases} 1 & x \in [0, 1) \\ 0 & \text{otherwise}. \end{cases}$$

Then the degree $p$ spline $\phi_p$ is defined as the $p + 1$ convolution of $\phi_0$:

$$\phi_p(x) = (\phi_0 * \phi_0 * \cdots * \phi_0)(x)$$

where $(f * g)(t) = \int f(t-s)g(s) \, ds$ denotes the convolution operator. In this paper, we will only consider splines of degree from $p = 0$ to $p = 2$ due to their overall computational simplicity. It can be easily deduced from (4) that the linear ($p = 1$) and quadratic ($p = 2$) cardinal splines are given as

$$\phi_1(x) = \begin{cases} x & x \in [0, 1) \\ 2 - x & x \in [1, 2) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\phi_2(x) = \begin{cases} \frac{1}{2} x^2 & x \in [0, 1) \\ \frac{1}{2} + (x - 1) - (x - 1)^2 & x \in [1, 2) \\ \frac{1}{2} - (x - 2) + \frac{1}{2}(x - 2)^2 & x \in [2, 3] \\ 0 & \text{otherwise} \end{cases}$$

respectively. Cardinal splines of degree $p$ have the appealing properties that (1) they are $p$ times differentiable and (2) are compactly supported with supp $\phi_p = (0, p + 1)$. Fig. 2 depicts the first three degree cardinal splines.

III. SPLINE SKETCH

In this paper, we are interested in constructing a new sketch that replaces the sinusoidal feature functions in the Fourier

$^2$The original Fourier sketch is a complex number, however it is possible to exploit the complex structure to simultaneously calculate sin and cos within FPGA implementations.
sketch with a set of scaled and translated spline functions. The sketch will consist of several splines that together cover the whole support of the timing depth window $T$, which we will treat as having periodic boundary conditions, exactly as in the Fourier setting. To do so, let $\xi_0, \xi_1, \ldots, \xi_M$ be a sequence of equispaced points, called knot points, over the whole timing depth window $T$ such that $\xi_0 = 0$ and $\xi_M = T$. In addition, let $\Delta := T/M$ denote the knot interval distance between two consecutive knot points. Then the scaled and translated spline of degree $p$ is defined as

$$\phi_{i,p}(x) = \phi_p \left( (x \mod T)/\Delta - i \right),$$

for $i = 0, 1, \ldots, M - 1$. The spline $\phi_{i,p}$ is supported on $[i\Delta, (i + p + 1)\Delta)$. As a result, a spline sketch of size $M$ can be expressed as

$$\hat{z} = \frac{1}{n} \sum_{i=1}^{n} \Phi_p(x_i)$$

where $\Phi_p(x) = [\phi_{i,p}(x)]_{i=0}^{M-1}$ is the set of spline feature functions of degree $p$. Fig. 3 illustrates the spline function feature $\Phi_p$ of size $M = 4$ for $p = 0, 1, 2$ as well as the real and imaginary components of the original Fourier sketch.

One can see straightforwardly that for $p = 0$ the spline feature function is equivalent to forming $M$ coarse bins over the whole timing depth window. As the degree $p$ increases, we observe two important points 1) the support of each spline feature increases linearly with respect to $p$, 2) each knot interval $\Delta$ contains $p + 1$ spline features. These two observations will become important when we consider the on-chip implementations of the spline sketches in Section III-A. Each of the Fourier features ($\phi_{\omega_j}(x) = e^{i\omega_j x}$) are depicted in the bottom right of Fig. 3. In contrast to the spline features, each individual Fourier feature has a support equal to the whole timing depth window $T$.

A. Hardware Implementation

Algorithm 1 details a prototype on-chip algorithm for computing a spline sketch of degree $p$ in an online manner. Once the appropriate knot interval of the photon detection has been established, only $p + 1$ of the total $M$ sketch features need to be updated. We detail further the total number of arithmetic operations required per photon detection for each of the $p = 0, 1, 2$ spline sketches.

For efficiency it makes sense to set $T$ and $M$ to integer powers of 2, and define $s := \log_2 T - \log_2 M$, so that identifying the knot interval can be read from the $\log_2 M$ leading bits of the $\log_2 T$-bit digital representation of the detection time. The remaining $s$-bits then represent the relative position within the knot interval which we denote by $r$. We further note that the cost of identifying the knot interval is the same for $p = 0, 1$ and 2. Therefore in examining the different arithmetic operations, we will only focus on the calculation of accumulated spline function values. For the spline function calculation, we assume that the spline functions are suitably scaled such that integer arithmetic can be used throughout. For the $p = 0$ degree spline function $\Phi_0$ (i.e. coarse binning), once we have located the appropriate bin all that is required is to update the bin count which involves a single one bit addition, e.g. implemented via a ripple counter.

In the case of the $p = 1$, two neighboring sketch values have to be calculated. From (6) we can see the first value is simply $r$ and requires no further calculation. The second value is $2^{s+1} - r$ and requires a single addition. These values then have to be added to the current sketch values requiring a further two additions. Therefore the linear spline needs to execute a total of 3 addition/subtractions per photon detection.

For $p = 2$, 3 spline values must be calculated per photon detection. In terms of the relative knot interval position, $r$, the 3 values that require calculation are: $r^2$, $2^s r - r^2$, and $r^2$. The dotted line sketches will consist of several splines that together cover the whole timing depth window.

Algorithm 1: On-Chip Sketch Processing.

Initialisation: Degree $p$, $\hat{z} = 0$, $n = 0$

While Acquisition Window do

If New Photon Arrival with Time Stamp $x$ then

Detect Knot interval $x \in [i\Delta, (i + 1)\Delta)$

For $q = 0$ to $p$ do

$\hat{z}_{i-q} = \hat{z}_{i-q} + \phi_{i-q,p}(x)$

End for

$n \leftarrow n + 1$

End if

End while

$\hat{z} \leftarrow \hat{z}/n$

Output: The empirical sketch $\hat{z}$ is transferred off-chip for post-processing.
$2^s - 2^{s+1}r + s^2$. Notice here that all scaling involved is by powers of 2 and therefore only involves a binary shift (whose computation we ignore). The resulting computation required is therefore 1 multiplication and 4 additions. Once calculated, these values will need to be added to the associated 3 sketch components (requiring 3 further additions). Therefore the quadratic spline needs to execute a total of 1 multiplication and 7 additions/subtractions per photon detection. In contrast, while there are various ways to compute the original Fourier sketch in (2), e.g. CORDIC or local polynomial approximation [14], depending on the latency and/or the computation desired, the computational cost is typically much higher. For example, calculating a Fourier sketch using a $w$-bit pipelined CORDIC algorithm would require more than $3Mw/2$ $w$-bit additions [14], roughly equivalent to $3M/2$ basic $w$-bit multiplications. Furthermore, in each case all $M$ sketch values need to be updated per photon detection. Specific details of the arithmetic operations required for each spline sketch per photon detection are summarized in Table I.

### IV. SPLINE SKETCH RECONSTRUCTION

In this section we show that spline sketches are amenable to simple piecewise reconstruction, including closed form solutions, multiple surface detection, and range walk correction. When considering piecewise reconstruction, we are naturally targeting a moderate range of SBR with a reasonable number of total photon detections per pixel (see Section V for details). While time-of-flight imaging is possible at very-low SBR and total photon detections, e.g. [16] (SBR $\sim$ 0.04 with only 2–3 photons detected per pixel) such methods invariably require the exploitation of spatial correlations across pixels and is not the focus of this current paper (see [8] for a sketching solution that exploits spatial information).

#### A. Closed Form Solution

We begin by proposing a simple closed form solution for the piecewise linear spline sketch$^3$ ($p = 1$) that accurately estimates the parameters of the observation model in (1) when we have a single isolated reflector with probability of detection $\alpha_1$, and probability of background detection, $\alpha_0 = (1 - \alpha_1)$. The method further assumes an IRF with local support and utilises adjacent linear splines located in the region of the reflected pulse to form a local mean estimator. Due to the semi-local nature that is inherent to splines (see Section III) one can easily establish a subset of splines that are needed to construct the local mean estimator. First we locate the individual spline sketch component

\[ \hat{\ell} = \arg \max_{1 \leq j \leq M} \hat{z}_j \]  

with the maximum magnitude,

\[ \hat{\ell} = \arg \max_{1 \leq j \leq M} \hat{z}_j \]  

where $\hat{z}_j$ is the $j$th entry of the sketch $\hat{z}$. Given that the largest individual sketch $\hat{z}_{\ell}$ has been located and assuming that the support of the IRF is less than one knot interval, then there are only 3 possible scenarios that can exist. Fig. 4 depicts each of the 3 possible scenarios.

In scenario 1, the IRF is contained within 1 kn interval on the left side of the $\ell$th spline. This is reflected in the bar chart (middle) of the spline sketch values. Similarly, in scenario 2, the IRF is contained within 1 kn interval on the right side of the $\ell$th spline. This is also reflected in the bar chart of the spline sketch values where the second largest sketch value is to the right of $\hat{z}_{\ell}$. Finally in scenario 3, the IRF straddles a knot point and is contained within both knot intervals, as captured by the bar chart of the spline sketch values. In each scenario, we can build a simple local mean estimator by taking the difference of adjacent sketch values about $\hat{z}_\ell$ as is captured in the third column in Fig. 4. Notice that when the IRF straddles a knot point, the local mean estimator is calculated by the difference between $\hat{z}_{\ell-1}$ and $\hat{z}_{\ell+1}$ hence the estimator is linear over 2 kn intervals. Specifically, the 3 estimators are calculated as follows

**Scenario 1:**  
\[ \hat{t} = \xi_{\ell} + \Delta \frac{z_{\ell} - z_{\ell-1}}{2M\alpha_1} \]  

**Scenario 2:**  
\[ \hat{t} = \xi_{\ell+1} + \Delta \frac{z_{\ell+1} - z_{\ell}}{2M\alpha_1} \]  

**Scenario 3:**  
\[ \hat{t} = \xi_{\ell+1} + \frac{z_{\ell+1} - z_{\ell-1}}{M\alpha_1} \]

In each case, the depth estimator $\hat{t}$ is dependent on knowing or accurately estimating $\alpha_1$. However, this can easily be estimated from neighbouring spline values that are assumed to only contain background photons. For the features located in positions that consist solely of photon detections from ambient sources it can be easily seen (see (1)) that

\[ \mathbb{P} \sum_{x, \ell} \phi_{\ell, p}(x) = \frac{\alpha_0}{M} = \frac{1 - \alpha_1}{M}. \]  

---

$^3$The same principle can be applied to $p = 2$ splines and is explored further in Section IV-C for range walk correction.

---

**TABLE I - ARITHMETIC OPERATIONS PER PHOTON DETECTION**

| Feature Function | Active Features | Add/Sub | Mult. |
|------------------|----------------|---------|-------|
| $\Phi_0$         | 2              | 0       | 1     |
| $\Phi_1$         | 2              | 3       | 0     |
| $\Phi_2$         | 3              | 7       | 1     |

---

**Fig. 4.** Left column: the 3 (out of 8) spline sketches in the region of the signal. Middle column: a bar chart of the spline sketches. Right column: local means estimator of each scenario.
Algorithm 2: Spline Sketch Matching Pursuit Reconstruction Algorithm.

Initialisation: degree p, number of reflectors K, \( \Theta = \emptyset \), residual \( r = \hat{z} \)

\( \text{for } k \leftarrow 1 \text{ to } K \text{ do} \)

Find a reflector highly correlated with the residual
\( \hat{t}_k \leftarrow \max_x \left\{ \frac{\mathbb{E}_{x \sim \pi}(x) \Phi_p(x) \cdot r}{\|\mathbb{E}_{x \sim \pi}(x) \Phi_p(x)\|_2} \right\} \)

Expected sketch value for the \( k \)th reflector
\( \hat{z}_k \leftarrow \mathbb{E}_{x \sim \pi}(x) \Phi_p(x) \)

\( \hat{\alpha}_k \leftarrow \frac{\|r \cdot \hat{z}_k\|}{\|\hat{z}_k\|^2} \)

\( \Theta \leftarrow \Theta \cup \{\hat{\alpha}_k, \hat{t}_k\} \)

\( r \leftarrow r - \hat{\alpha}_k \hat{z}_k \)

end for

Normalise \( \alpha \) such that \( \sum_{k=0}^{K} \alpha_k = 1 \).

Output: Estimates for the depth and intensity information of \( K \) reflectors in a given pixel.

One can therefore identify a set \( B \) of sketch values \( \hat{z}_j \) in the vicinity of the maximum that are known to only contain background counts, and use these to estimate \( \alpha_1 \) as follows:

\[
\alpha_1 = 1 - \frac{M}{|B|} \sum_{i \in B} \hat{z}_i. 
\]

(15)

Given the 3 estimators above, one can choose the estimate which produces the lowest spline sketch loss:

\[
\|\hat{z} - \mathbb{E}_{x \sim \pi} \Phi_1(x)\|_2. 
\]

(16)

As one has to only compute the loss in (16) a total of 3 times, the computational complexity of this closed form solution is minimal. As the method returns the (local) mean of the IRF distribution, it is also independent of the IRF shape (incorporating the necessary leading-edge to mean correction as appropriate). For additional accuracy one could further fine-tune the estimates by running a few iterations of a spline version of the SMLE algorithm proposed in [6] or additionally incorporate powerful spatial denoisers as proposed in [8]. We evaluate the performance of this closed formed estimator in Section VI and compare the error it produces to the optimal Cramér-Rao lower bound.

C. Range Walk Correction

Lidar systems are subject to systematic range errors that arise due to a difference between the estimated results and the ground truth distance of the object in the scene [19], [20]. This can happen due to many aspects of the detector’s response, but it is particularly linked to imaging highly reflective objects such as a retroreflector. Retroreflectors are ultra reflective surfaces that lead to a high probability of photon detection per laser pulse. Due to the intrinsic deadtimes of the lidar device, the SPAD detection deviates from the classical Poisson statistical model and more sophisticated models need to be considered [20]. In the medium and large flux regimes, the SPAD tends to detect photon arrivals at earlier bins along the timing depth window, resulting in an observed IRF that is distorted towards the leading edge of the original pulse. This phenomena is commonly referred to as photon pile-up and can cause large range errors for highly reflective surfaces.

Fig. 5 depicts the distortion of a real life IRF from the face dataset in [22] simulated using the multinomial distribution in [20] for different target reflectivities (see Appendix A for more details). It can be seen that as the target reflectivity increases,

\( 4 \) Photon pile-up can also occur due to high background illumination (e.g. strong sunlight). We do not consider this case here as it is usually mitigated through the use of multiple SPADS per pixel and coincidence detection, see, e.g., [19], [21].
As the location of the target reflectivity increases, the observed IRF becomes as a linear combination of quadratic spline functions. The right of Fig. 6 shows the range walk error as a function of the local estimate of the standard deviation which can be calculated using the quadratic spline sketch (here \( M = 50 \)) following the same principle as the local mean estimators presented in Section IV-A, as detailed below. In contrast to the intensity based range walk model, a shape based range walk model can define a robust local look-up table correction over the full range of reflectivities.

1) A Local Closed Form Standard Deviation Estimate: In a similar fashion to Section IV-A, one can construct a local standard deviation estimator in closed form that is both simple and computationally cheap. First we locate the individual spline sketch with the maximum magnitude:

\[
\ell = \arg \max_{1 \leq j \leq M} \hat{z}_j
\]

where \( \hat{z}_j \) is the \( j \)th entry of the quadratic spline sketch. Next, we remove the noise contribution from the individual sketches by defining the noise-corrected sketches:

\[
\tilde{z}_j = \frac{1}{\alpha_1} \left( \hat{z}_j - \frac{1}{M} \right),
\]

for \( j = 1, \ldots, M \), where \( \alpha_1 \) can be estimated as before using (15).

Given the local neighbourhood of 5 noise-corrected sketch values around \( \hat{z}_{\ell} \):

\[
\tilde{z}_{\text{sub}} = [\tilde{z}_{\ell-2}, \tilde{z}_{\ell-1}, \tilde{z}_{\ell}, \tilde{z}_{\ell+1}, \tilde{z}_{\ell+2}]^\top,
\]

we can calculate the local first and second moments of the observed IRF by using the quadratic splines to compute a local linear function, \( f_1(x) = \sum_{i=1}^{N} c_1(i) \phi_{\ell+i-3,2}(x) \), and a local quadratic function, \( f_2(x) = \sum_{i=1}^{N} c_2(i) \phi_{\ell+i-3,2}(x) \), as illustrated in Fig. 7. The coefficients \( c_1 \) and \( c_2 \), with a little algebra, can be shown to be given by:

\[
c_1 = [-2\Delta, -\Delta, 0, \Delta, 2\Delta]^\top,
\]

\[
c_2 = \left[ \frac{15}{4} \Delta^2, \frac{3}{4} \Delta^2, \frac{3}{4} \Delta^2, -\frac{1}{4} \Delta^2, \frac{15}{4} \Delta^2 \right]^\top
\]

recalling that \( \Delta = \frac{T}{M} \).

It can then easily be shown that the depth and standard deviation estimates have the following closed form solutions

\[
\hat{\xi}_\ell = \xi_{\ell+1} + \frac{\Delta}{2} + c_1^\top \tilde{z}_{\text{sub}},
\]

and

\[
\hat{\sigma}^2 = c_2^\top \tilde{z}_{\text{sub}} - \left( c_1^\top \tilde{z}_{\text{sub}} \right)^2.
\]

---

5The intensity of a single target is defined as the number of recorded photons associated with the target, i.e., \( \eta \alpha \).
For further details we refer the reader to the MATLAB code examples whose link can be found in Section VII.

V. CRAMÉR-RAO ERROR BOUNDS

In this section, the Cramér-Rao error bounds of the spline sketches for $p = 0, 1, 2$ are compared with the original Fourier sketch as well as the full TCSPC histogram (i.e. no compression). The Cramér-Rao bound (CRB) gives a lower bound for the root mean squared error (RMSE) of an estimator $\hat{\theta}$ and can therefore provide the best case performance one can achieve from a specific sketch, independent of the reconstruction algorithm.

Given the observation model $\pi$ in (1) and the corresponding Fisher information matrix (FIM), defined as

$$I_{\text{data}}(\theta) = nE_{x \sim p} \left[ \frac{\partial \log \pi(x | \theta)}{\partial \theta} \frac{\partial \log \pi(x | \theta)}{\partial \theta}^\top \right],$$

then the optimal Cramér-Rao RMSE, in terms of the full data, is defined as

$$\text{RMSE}_n = \sqrt{\sum_{k=1}^{2K} [I_{\text{data}}(\theta)^{-1}]_{kk}}.$$  \hspace{1cm} (23)

Equivalently, for the sketched case the FIM is defined as (see for instance [6])

$$(I_{\text{sketch}}(\theta))_{ij} = nE_{\Phi} \left[ \frac{\partial z_\theta}{\partial i} \frac{\partial z_\theta}{\partial j} \right],$$

where as before $z_\theta = E_{x \sim \phi} \Phi_p(x)$ and where $\Sigma_\theta = \text{Cov}(z_\theta)$ denotes the $M \times M$ covariance matrix of the spline sketch. Similarly, we can define the optimal sketched Cramér-Rao RMSE as

$$\text{RMSE}_M = \sqrt{\sum_{k=1}^{2K} [I_{\text{sketch}}(\theta)^{-1}]_{kk}}.$$  \hspace{1cm} (25)

A. Factors Affecting the CRB

The CRB is subtly characterised by a number of factors which we now discuss. From the observation model (1) the underlying difficulty in parameter estimation depends on: the probabilities of detection for the reflectors, $\alpha_1, \ldots, \alpha_K$, and the background, $\alpha_0$, the total number of photons detected, $n$, within $N$ laser pulses, the pulse width, $\sigma$, and the size of the timing depth window, $T$. For a given sketch size $M$ and a single reflector, the CRB is therefore a function of $\{M, n, SBR, \sigma, T\}$. It is important to note that SBR is not a signal-to-noise measure. Indeed, as can be seen from (24), for a fixed SBR the CRB can be made arbitrarily small by simply increasing the number of laser pulses and hence the total number of photons detected [16]. Similarly, for fixed SBR, decreasing the pulse width reduces the uncompromised CRB. In contrast, assuming a fixed background illumination (i.e. $\alpha_0/T = \text{const.}$), increasing $T$ decreases the SBR without significantly changing the CRB. Finally, we note that it is also possible to artificially boost $n$ through spatial regularisation, e.g. [8], [16], to achieve good performance at very low SBR and total number of photons detected. However, we do not consider such models here.

For simplicity we assume that the IRF is a Gaussian function with pulse width $\sigma$. This ensures that the formula in (25) has a closed form solution, however, one can expect similar results to hold for more general IRFs. We keep the size of the timing depth window fixed at $T = 600$ bins (equivalent to a depth range of 24 m) and the sketch size fixed at $M = 8$, while varying the SBR and pulse width. We also highlight the difference between the CRB as a function of depth and averaged over depth.

B. Spatially Averaged CRB Performance

We begin by exploring the spatially averaged CRB performance of the different sketches. Each result is averaged along 1000 uniform depths across the timing depth window $[0, T = 1]$. In our first experiment, we vary the SBR from a low ($10^{-1}$) to a high ($10^1$) SBR regime. Here the acquisition time corresponds to a depth range 24 m where each of the $T$ timing bins equates to a precision of 4 cm. The sketch size for each of the spline and Fourier sketches is fixed at $M = 8$ and the number of photon detections is set to $n = 1000$. Fig. 8 shows the CRB for the different sketches for a Gaussian pulse width of $\sigma = 64$ cm for spline sketches of degree $p = 0, 1, 2$, a Fourier sketch and a full data TCSPC histogram (i.e. no compression). First, while the degree $p = 0$ spline sketch (i.e. coarse binning) is able to achieve good performance at high SBR this reduces as the SBR gets less. Furthermore the $p = 0$ RMSE is the highest of all the sketches. There is a significant reduction of RMSE as the degree of the spline is increased to $p = 1$. Notably, the improvement of using a degree $p = 2$ spline sketch is negligible in comparison to its linear counterpart. This is potentially due to the fact that locally first order statistics are almost sufficient in capturing the depth information and therefore the RMSE saturates for $p \geq 1$. There is a slight loss of information in comparison to using the original Fourier sketch of the same size, however one gains in the computational simplicity of constructing the spline sketches. Nonetheless, for a moderate SBR of 1, the RMSE of all the estimators apart from coarse binning are approximately the same at 5 cm.

Fig. 9 shows the results for the same set up as before but now with a narrower Gaussian pulse width of $\sigma = 24$ cm, as might occur in medium-high flux scenarios. In this case, one can clearly see that the coarse binning method fails catastrophically even at
SHEEHAN et al.: SPLINE SKETCHES: AN EFFICIENT APPROACH FOR PHOTON COUNTING LIDAR

Fig. 9. CRB for varying SBR with a fixed pulse width $\sigma = 24$ cm, total number of photons, $n = 1000$, and sketch size, $M = 8$.

Fig. 10. CRB for varying pulse width with fixed SBR = 1, total number of photons, $n = 1000$, and sketch size, $M = 8$.

Fig. 11. CRB for varying depth positions with fixed SBR = 1, pulse width, $\sigma = 64$ cm, total number of photons, $n = 1000$, and sketch size, $M = 8$.

Fig. 12. A comparison of the proposed reconstruction algorithms and their respective CRBs varying the SBR with a fixed pulse width $\sigma = 64$ cm, total number of photons, $n = 1000$, and sketch size, $M = 8$.

Fig. 13. Average intensity RMSE as a function of the number of observed photons with SBR fixed at 0.1, pulse width, $\sigma = 24$ cm, and sketch size fixed at $M = 8$.

Fig. 12. A comparison of the proposed reconstruction algorithms and their respective CRBs varying the SBR with a fixed pulse width $\sigma = 64$ cm, total number of photons, $n = 1000$, and sketch size, $M = 8$.

Analyzing the poor performance of coarse binning with narrow IRFs reveals a more troubling phenomenon. The results presented in the previous subsection averaged the CRBs over the whole timing depth window. However, the CRBs, with the exception of the Fourier sketch, do vary as a function of target depth. Therefore the average CRBs may be hiding a substantially worse performance at certain depths. We therefore next investigate the spatial dependency of the CRBs for the different sketches. To do so, the SBR and pulse width are fixed to 1 and 64 cm, respectively, and the depth of the reflector is varied along the range of the lidar device. Fig. 11 shows the results.

C. CRB Spatial Dependence

Analyzing the poor performance of coarse binning with narrow IRFs reveals a more troubling phenomenon. The results presented in the previous subsection averaged the CRBs over the whole timing depth window. However, the CRBs, with the exception of the Fourier sketch, do vary as a function of target depth. Therefore the average CRBs may be hiding a substantially worse performance at certain depths. We therefore next investigate the spatial dependency of the CRBs for the different sketches. To do so, the SBR and pulse width are fixed to 1 and 64 cm, respectively, and the depth of the reflector is varied along the range of the lidar device. Fig. 11 shows the results. As expected, the CRBs for the full data and Fourier sketch are independent of the depth of the reflector. In contrast, the coarse binning estimator ($p = 0$) varies from a small RMSE at the knot point to a significantly larger maximum error ($\sim 28$ cm) at the centre of the knot interval. This can be intuitively understood as follows. If the return pulse straddles a knot point then its high SBR levels. The reason for this will become clear in the next subsection. In contrast, the other methods achieve similar performance to before and in fact are slightly more accurate. In both Figs. 8 and 9, we see that each of the $p = 1, 2$ degree splines and the Fourier sketches converge to the device’s resolution (4 cm in this case) as the SBR becomes high.

To further understand the relationship between the pulse width and the RMSE of the estimators, we fix the SBR at 1 and vary only the pulse width within the range of 2 and 100 cm. Fig. 10 demonstrates that the coarse binning RMSE is extremely dependent on the pulse width of the IRF, whilst, in contrast, the other methods are agnostic to the pulse shape. Some practitioners, for instance [5], have proposed increasing the pulse width of the device’s laser to achieve a coarse binning RMSE that is only slightly worse than using either the sketches or full data. However, this is only practical in certain circumstances, and if the device experiences photon pile-up (see Section IV-C) then the width of the detected signal can collapse towards a Dirac function. In such an event, coarse binning would fail catastrophically as illustrated in Fig. 10.
precise position can be estimated accurately from the relative magnitudes of the neighboring histogram bins. However, if the pulse lies solely in one knot interval then the estimator is blind to where precisely the pulse is within the knot interval. In contrast, the spline sketches for degrees $p = 1, 2$ exhibit minimal spatial dependency. Interestingly, where the $p = 1$ spline sketch achieves the largest RMSE is at the knot points and achieves the smallest RMSE at the centre of the knot intervals which is the reverse to that for the even degree splines sketches of $p = 0, 2$. This result importantly highlights the worst case performance one can expect to achieve from each spline sketch which, as demonstrated, can be extreme for coarse binning yet is controllable for both the linear and quadratic spline sketches.

VI. Simulations

In this section we evaluate the proposed algorithms on both synthetic and real datasets and compare their performance to the CRBs discussed in Section V.

A. Algorithm Dependent Performance

While the CRBs provide us with performance bounds for any estimator using the associated statistics, it is also necessary to understand how a given algorithm performs in practice. We therefore begin this section by showing how the different algorithms presented in Section IV compare against the CRBs for depth estimation. The set up for this experiment is identical to the one presented in Fig. 8. We vary the SBR from a $10^{-1}$ to 10. The depth range is fixed at 24 m with $T$ timing bins equating to a precision of 4 cm each. The sketch size is $M = 8$, the total number of photon detections is $n = 1000$, and the IRF is modelled as a Gaussian pulse with width $\sigma = 64$ cm. Fig. 12 shows that all algorithms perform equivalently at high SBR. While for low SBR all algorithms exhibit a small drop in performance compared with their associated CRB. The MP algorithm appears to perform similarly for both $p = 1$ and $p = 2$ while the closed form estimate is slightly worse. This is probably due to the fact that the Gaussian IRF is not fully supported within a single knot interval width as was assumed in the closed form derivation and could likely be remedied by additional fine-tuning as discussed in Section IV-A.

While the most challenging parameter to estimate accurately from compressed statistics is the depth, for completion, we next consider the accuracy of the intensity reconstruction for the different algorithms. Fig. 13 shows the intensity error in a low SBR regime, SBR $= 0.1$, as a function of total photon count, $n$. All other parameters are as above. Here, we see that all methods perform well at moderate values of $n$. However, as $n$ reduces below 100 total photons detected, the error for the sketching and coarse histogramming methods increases more...
for the uncompressed matched filter, with the histogramming method performing worst of all.

**B. Polystyrene Head Test Data**

We next evaluate the proposed spline sketches and algorithms in Section IV on a real dataset consisting of lidar returns from a polystyrene head placed at a distance of 40 metres [22]. We compare the proposed methods with the cross correlation algorithm [1] that works directly on the full TCSPC (i.e., no compression), as well as the SMLE algorithm in [6] using Fourier features to construct the sketch. The polystyrene head dataset has size of $141 \times 141$ pixels with $T = 4613$. Most of the pixels in this scene contain exactly $K = 1$ surface and the average number of photon detections per pixel is 337. Fig. 14 and Table II show the reconstruction of the scene and the corresponding RMSE metrics, respectively, using a Fourier sketch and the SMLE algorithm in [6], coarse binning ($p = 0$), the closed form solution in Section IV-A, and the $p = 1, 2$ spline sketches using the MP algorithm in Section IV. These compression algorithms are compared to the matched filtering algorithm [1] that acts on the original uncompressed TCSPC histogram. For the compression techniques, we consider a sketch size of $M = 10, 20, 30$ and 40. Initially, we see that the coarse binning method fails to accurately recover the scene for the smaller sketch sizes. In comparison, both $p = 1, 2$ degree splines retain enough information from the ToF data to allow for accurate reconstruction. One can observe a slight drop in performance between the closed form solution and the matching pursuit algorithm especially around the border of the head. Interestingly, a linear or quadratic spline sketch of just $M = 20$ allows for very accurate reconstruction in comparison to both the Fourier sketch and the full data reconstruction.

Table II also shows storage costs (bits per pixel) and digital logic requirements (number of logic gates per pixel) for the different sketching schemes for this experiment. The storage costs are based on the assumption that the maximum number of bits needed to store a spline sketch is $6p + s \log_2 n$, while the number of logic gates beyond a basic histogram construction is calculated based on standard ripple adder and combinational

---

**TABLE II**

| Sketch Size $M$ | TCSPC gpp | TCSPC bpp | Coarse binning $p = 0$ gpp | Coarse binning $p = 0$ bpp | $p = 1$ gpp | $p = 1$ bpp | $p = 2$ gpp | $p = 2$ bpp | Fourier gpp | Fourier bpp |
|-----------------|------------|-----------|-----------------------------|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 10              | 100        | 4.4       | 100                         | 126                         | 190         | 12.1        | 2412        | 280         | 11.7        | 23490       |
| 20              | 200        | 111       | 200                         | 360                         | 8.4         | 1950        | 520         | 8.5         | 36960       | 520         |
| 30              | 300        | 111       | 300                         | 340                         | 6.2         | 1737        | 780         | 6.4         | 48600       | 780         |
| 40              | 400        | 96        | 400                         | 680                         | 5.7         | 1336        | 960         | 5.9         | 56240       | 960         |

*The matched pursuit algorithm was used for the sketching solutions of $p = 1$ and $p = 2$.**

---

For a spline of order $p$, we have up to $2^{ps}$ different values per received photon where $s = \log_2 T - \log_2 M$, and we can record up to $n$ total photons.
multiplier gate counts and the arithmetic complexity derived in Section III. Here, we note that for \( p = 1 \) we need to process word lengths of \( s \)-bits while for \( p = 2 \) word lengths of \( 2s \)-bits are needed. For the Fourier sketch we have assumed a CORDIC implementation \([14]\) with the same word length, \( 2s \)-bits, as for the \( p = 2 \) spline sketch. As can be seen from the table, the required logic is modest for the spline sketches with the quadratic spline costing roughly 20 times more than for the linear spline. In both cases the cost decreases for larger sketch sizes (due to the reducing size of \( s \)). In contrast, the Fourier sketch is over an order of magnitude larger, with the cost growing roughly linearly with sketch size. For a given sketch size, \( M \), the storage cost of the higher order splines is slightly greater than for the coarse histogram. However, this is more than compensated for by the dramatic improvements in RMSE as shown in Fig. 14. For this specific dataset, the linear and quadratic spline sketches correspond to compression of approximately 95% in comparison to storing the raw TCSPC photon detection times without sacrificing the overall quality of the reconstructed image.

\section{C. Range Walk Correction Simulations}

To evaluate our shape-based range walk model discussed in Section IV-C, we use the IRF in \([22]\) where \( T = 4613 \) (see Section VI-B for more details) and build a look-up table as in Fig. 6 for a sketch size of \( M = 25, 50, \ldots, 125 \) using the range walk model in Appendix A. Once an intensity or standard deviation value has been estimated from the spline sketches, we consult the aforementioned look-up tables using the estimates to correct for the range walk. For 50 different target reflectivities (on a log-scale between \( 10^{-4} \) and 1) and SBR values of 100, 10, 1 and 0.1, we compute the root mean squared error (RMSE) calculated over 250 random target distances. Fig. 15 shows the RMSE for each of the SBR values as a function of target reflectivity. One can initially see that as the point of intensity saturation is hit, the RMSE grows rapidly as the intensity based range walk model is unable to accurately distinguish the correct range walk. In contrast, the shape based range walk model stays at quite a constant error level throughout despite the increase in target reflectivity. For the lowest SBR of 0.1, the smaller sketch sizes of \( M = 25 \) and \( M = 50 \) have a larger RMSE for the higher target reflectivity. This is because in challenging imaging scenes, a larger sketch size is required to accurately estimate and remove the contribution from background sources. These results demonstrate additional statistical information captured by the quadratic spline sketch provide further benefits in complex environments.

\section{VII. Conclusion}

In this paper, we have proposed a semi-local spline sketch that can be efficiently implemented in digital logic and that achieves the compression capabilities of the original Fourier sketch whilst enjoying minimal overhead computations on-chip. The simplest \( p = 1 \) spline sketch can be computed without any hardware multiplication and admits a simple closed form reconstruction. Increasing the degree of the spline to \( p = 2 \) requires a modest increase in computation but further enabled us to capture local 2nd order statistics of the return pulses that allow us to accurately correct for range walk error where intensity based range walk models fail to do so. Simulations demonstrated that both the linear and quadratic splines sketches achieve almost the same performance as the original Fourier sketch with only a slight drop in performance for very low SBR scenes. The next steps in this research are to develop a full digital logic implementations of the sketches for complete hardware evaluation.

\section{Code Availability}

A MATLAB implementation of all the algorithms discussed are available at the repository https://gitlab.com/mpsheehan1995/spline-sketch-lidar.

\section{Appendix}

\subsection{Range Walk Simulation}

The IRFs illustrated in Fig. 5 and used in the experiments in Section IV-C were generated using the pile-up model in \([20]\). The probability of detecting a photon in the \( m \)th laser repetition out of \( N \) is

\begin{equation}
\pi_0 = 1 - \exp \left( - \sum_{i=0}^{T-1} \lambda_i \right)
\end{equation}

where \( \lambda_i \) is the photon flux on the SPAD during the \( m \)th time bin of the timing depth window defined as

\begin{equation}
\lambda_i = \sum_{k=1}^{K} \frac{\mu_b \beta_k}{H} h(i - t_k) + \mu b
\end{equation}

with \( \mu \) the SPAD photon detection probability (\( \sim 1\% \)), \( \beta_k \in [0, 1] \) the \( k \)th target reflectivity and \( b \) the level of detections from background sources. If a photon is detected, its time stamp follows the distribution

\begin{equation}
\pi_{rw}(x|\theta) = \frac{1}{\pi_0} \left( \exp \left( - \sum_{i=0}^{T-1} \lambda_i \right) - \exp \left( - \sum_{i=0}^{T-1} \lambda_i \right) \right)
\end{equation}

Note that for low flux \( \lambda_i \ll 1 \), this distribution is well approximated by the simpler model in (1) with

\begin{equation}
\alpha_0 = \frac{bT}{\sum_{k=1}^{K} \beta_k + bT}
\end{equation}

and

\begin{equation}
\alpha_k = \frac{\beta_k}{\sum_{k=1}^{K} \beta_k + bT}
\end{equation}

for \( k = 1, \ldots, K \).

\section{References}

[1] A. McCarthy, R. J. Collins, N. J. Krichel, Y. Fernández, A. M. Wallace, and G. S. Buller, “Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting,” Appl. Opt., vol. 48, no. 32, pp. 6241–6251, Nov. 2009.
[2] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, “Single-photon three-dimensional imaging at up to 10 kilometers range,” Opt. Exp., vol. 25, no. 10, pp. 11919–11931, May 2017.
[3] R. K. Henderson et al., “A 192 × 128 time correlated single photon counting imager in 40nm CMOS technology,” in Proc. IEEE 44th Eur. Solid State Circuits Conf., 2018, pp. 54–57.
[4] S. W. Hutchings et al., “A reconfigurable 3-D stacked SPAD imager with in-pixel histogramming for flash LIDAR or high-speed time-of-flight imaging,” *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, Nov. 2019.

[5] J. Gyongy et al., “High-speed 3D sensing via hybrid-mode imaging and guided upsampling,” *Optica*, vol. 7, no. 10, pp. 1253–1260, 2020.

[6] M. P. Sheehan, J. Tachella, and M. E. Davies, “A sketching framework for reduced data transfer in photon counting lidar,” *IEEE Trans. Comput. Imag.*, vol. 7, pp. 989–1004, 2021.

[7] M. P. Sheehan, J. Tachella, and M. E. Davies, “Surface detection for sketched single photon lidar,” in *Proc. IEEE 29th Eur. Signal Process. Conf.*, 2021, pp. 621–625.

[8] J. Tachella, M. P. Sheehan, and M. E. Davies, “Sketched RT3D: How to reconstruct billions of photons per second,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 1566–1570.

[9] M. P. Sheehan, J. Tachella, and M. E. Davies, “Hardware friendly spline sketched lidar,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2023, pp. 1–5.

[10] Y. Altman and S. McLaughlin, “Range estimation from single-photon lidar data using a stochastic em approach,” in *Proc. Eur. Signal Process. Conf.*, 2018, pp. 1112–1116.

[11] P. Padmanabhan, C. Zhang, and E. Charbon, “Modeling and analysis of a direct time-of-flight sensor architecture for lidar applications,” *Sensors*, vol. 19, no. 24, 2019, Art. no. 5464. [Online]. Available: https://www.mdpi.com/1424-8220/19/24/5464

[12] R. Gribonval, A. Chatelaz, N. Keriven, V. Schellekens, L. Jacques, and P. Schniter, “Sketching data sets for large-scale learning: Keeping only what you need,” *IEEE Signal Process. Mag.*, vol. 38, no. 5, pp. 12–36, Sep. 2021.

[13] A. Hall, *Generalized Method of Moments*. London, U.K.: Oxford Univ. Press, 2007, pp. 230–255.

[14] F. de Dinechin, M. Istoon, and G. Sergent, “Fixed-point trigonometric functions on FPGAs,” 2014. [Online]. Available: https://doi.org/10.1145/2641361.2641375

[15] M. Unser, “Splines: A perfect fit for signal and image processing,” *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[16] J. Rapp and V. K. Goyal, “A few photons among many: Unmixing signal and noise for photon-efficient active imaging,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 3, pp. 445–459, Sep. 2017.

[17] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, “Sketching for large-scale learning of mixture models,” *Inf. Inference, A J. IMA*, vol. 7, no. 3, pp. 447–508, 2018.

[18] A. H. Stroud, “Methods of numerical integration,” *SIAM Rev.*, vol. 28, no. 1, pp. 128–128, 1986. [Online]. Available: https://doi.org/10.1137/1028042

[19] A. Tontini, L. Gasparini, and M. Perenzoni, “Numerical model of SPAD-based direct time-of-flight flash lidar CMOS image sensors,” *Sensors*, vol. 20, no. 18, 2020, Art. no. 5203. [Online]. Available: https://www.mdpi.com/1424-8220/20/18/5203

[20] F. Heide, S. Diamond, D. B. Lindell, and G. Wetzstein, “Sub-picoscend photon-efficient 3D imaging using single-photon sensors,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, 2018.

[21] A. Incoronato, M. Locatelli, and F. Zappa, “Statistical modelling of SPADS for time-of-flight lidar,” *Sensors*, vol. 21, no. 13, 2021, Art. no. 4481. [Online]. Available: https://www.mdpi.com/1424-8220/21/13/4481

[22] Y. Altman, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, “Lidar waveform-based analysis of depth images constructed using sparse single-photon data,” *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1935–1946, May 2016.

[23] M. S. Oh, H. J. Kong, T. H. Kim, K. H. Hong, and B. W. Kim, “Reduction of range walk error in direct detection laser radar using a Geiger mode avalanche photodiode,” *Opt. Commun.*, vol. 283, no. 2, pp. 304–308, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401809009663

Michael P. Sheehan (Member IEEE) received the integrated master’s degree in mathematics from the University of Manchester, Manchester, U.K., in 2017, and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 2022. He was a Postdoctor with the University of Edinburgh till 2022. He is currently ML Scientist for LNER within the transport industry. His research interests include computational imaging, inverse problems, and efficient based machine learning.

Julían Tachella (Member IEEE) received the electronic engineering degree (Hons.) from Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina, in 2016, and the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., in 2020. From 2020 to 2021, he was a Postdoctor with the University of Edinburgh. He currently holds a Centre National de Recherche Scientifique research scientist position with École Normale Supérieure de Lyon, Lyon, France. His research interests include inverse problems, and deep learning, and applications in various computational imaging problems, such as single-photon lidar and non-line-of-sight imaging.

Mike E. Davies (Fellow, IEEE) received the M.A. degree in engineering from Cambridge University, Cambridge, U.K., and the Ph.D. degree in nonlinear dynamics and signal processing from University College London, London, U.K. He currently holds the Jeffrey Collins chair in signal and image processing with the University of Edinburgh, Edinburgh, U.K., where he is also the Director of Research with the School of Engineering. His research interests include low dimensional signal models, compressed sensing, computational imaging, and machine learning. He was the recipient of various awards and prizes including a Royal Society University Research Fellowship, an ERC Advanced Grant, and Royal Society Wolfson Research Merit Award. He is a Fellow of EURASIP, Royal Academy of Engineering, and Royal Society of Edinburgh.