




Asynchrony Increases Efficiency: Time Encoding of Videos and Low-Rank Signals

Karen Adam , Adam Scholefield , *Member, IEEE*, and Martin Vetterli , *Fellow, IEEE*

Abstract—In event-based sensing, many sensors independently and asynchronously emit events when there is a change in their input. Event-based sensing can present significant improvements in power efficiency when compared to traditional sampling, because (1) the output is a stream of events where the important information lies in the *timing* of the events, and (2) the sensor can easily be controlled to output information only when interesting activity occurs at the input. Moreover, event-based sampling can often provide better resolution than standard uniform sampling. Not only does this occur because individual event-based sensors have higher temporal resolution (Rebecq et al., 2021) it also occurs because the asynchrony of events within a sensor and therefore across sensors allows for less redundant and more informative encoding. We would like to explain how such curious results come about. To do so, we use ideal time encoding machines as a proxy for event-based sensors. We explore time encoding of signals with low rank structure, and apply the resulting theory to video. We then see how the *asynchronous* firing across time encoding machines can couple spatial sampling density with temporal resolution, leading to better reconstruction, whereas, in frame-based video, temporal resolution depends solely on the frame-rate and spatial resolution solely on the pixel grid used.

Index Terms—Event-based sensing, time encoding, low-rank signals, bandlimited signals, video reconstruction.

I. INTRODUCTION

MANY aspects of our lives are governed by routine and rhythm: our work days, breathing patterns, or even music. However, metronomic schedules aren't always resource-efficient. We generally say “hello” *when* we see someone we know rather than saying it at regular intervals.

Many engineered systems, such as traditional sampling devices, rely almost exclusively on clocked behavior. These sampling schemes are powerful - they govern how we record music, take images, transfer information - but they fail to adapt their activity to the varying complexity of the input.

This drawback leads to inefficiencies which are apparent when comparing the power consumption of human-engineered technologies to biological equivalents.

Manuscript received May 29, 2021; revised October 14, 2021 and November 18, 2021; accepted November 19, 2021. Date of publication December 9, 2021; date of current version December 31, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hassan Mansour. This work was supported by the Swiss National Science Foundation under Grant 200021_181978/1, SESAM - Sensing and Sampling: Theory and Algorithms. (*Corresponding author: Karen Adam.*)

The authors are with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: karen.adam@epfl.ch; adam.scholefield@epfl.ch; martin.vetterli@epfl.ch).

Digital Object Identifier 10.1109/TSP.2021.3133709

As demands for increased storage and processing coupled with smaller devices has brought efficiency into the spotlight, researchers have been turning to biology for inspiration.

Inspired by neurons, event-based sensing is growing in popularity [2]–[5]. The output of such a sensor is a series of spikes which are characterized by their *timing* rather than their amplitude, as is the case with traditional sampling [6]. In addition, the spiking rate at the output can be made to depend on different characteristics of the input, such as amplitude or rate of change, by choosing an appropriate filter [7].

While efforts have recently been invested to better understand event-based sensing and reconstruction of bandlimited or finite-rate-of-innovation signals [9]–[11], comparisons between event-based sensing and standard sampling have mostly considered the timing-based output of event based sensing to be more of a pesky necessity that requires some work-around rather than a blessing in disguise. In fact, this asynchrony between the events of different sensors allows for better resolution and more flexibility when using event-based sampling than in the uniform, synchronous sampling case. Having asynchrony between events of different sensors is one way of ensuring that the information gathered about an input signal provides linearly independent—and therefore non-redundant—constraints on this input.

We show this in a series of steps. First, we review time encoding machines (TEMs), which are the ideal event-based sensor [7], [12]. The input-output relationship of an integrate-and-fire TEM follow similar rules to that of nonuniform sampling [13], [14], and thus results in interesting consequences when it comes to single-signal multi-channel time encoding [15], [16].

The natural extension to time encoding multiple mixed signals with multiple channels then offers optimal sampling efficiency for low-dimensional signals with known structure. We find a Nyquist-like criterion on the number of spikes needed for reconstruction, requiring as many linearly independent constraints as degrees of freedom.

Using this formulation, the video recording problem with TEMs turns into a parametric estimation problem. We study the setup in Fig. 1: multiple TEMs are used to encode multiple locations in a scene and each TEM outputs a series of spikes. We find that this setup offers interesting tradeoffs in terms of time and space resolution: for once, increasing space resolution can actually *increase* time resolution as well, precisely thanks to our blessing in disguise—the asynchrony of spike times within and across outputs of the TEMs.

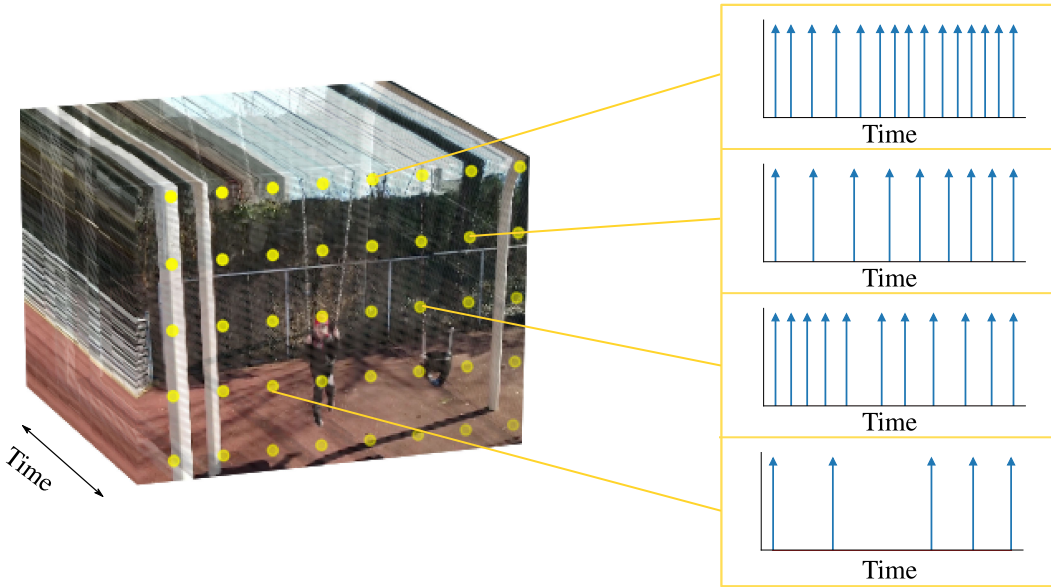


Fig. 1. Vision setup: we assume that we have an array of spiking devices, such as photoreceptors or TEMs, each of which is observing a scene at a particular location. The input to the receptor at this location is a time varying signal and the receptor will output a stream of spikes, the timing of which is dependent on the input. On the left, we show the projection of the scene which is being observed, with an overlay of event-based sensors shown in yellow. To its right, we zoom in to view the spiking output of some of the sensors. The video used is taken from the Need for Speed dataset [8].

As a result, time encoding or event-based sensing encourages increasing the number of sensors rather than the spiking rate of sensors as a means to improve resolution in both time and space. This, in turn, means that (1) there are fewer hardware requirements on the sensors themselves, and (2) sensors can integrate information over more time, thus avoiding issues with low photon count that occur at higher shutter speeds with standard cameras.

This particular result does not usually hold in the equivalent scenario in standard frame-based video recording. There, the resolution in time and space are almost independent from one another. The former depends on the frame-rate of the camera and the latter depends on the sampling pattern of the camera, which of course includes the number of pixels used for sampling. If we increase the number of pixels used to record a frame, this would not improve the resolution in time, because all pixels of a frame are taken at the same *moment* in time. To fix this, different pixels would need to record frames at *different* times. This is possible but renders things more complicated: reconstruction would either require that time shifts between pixel clocks be known, or it becomes a difficult problem that has no uniqueness guarantees. In contrast, when using time encoding, spike times are—by design—almost surely different and this difference comes at no extra cost.

II. BACKGROUND

As first presented in [17], time encoding machines (TEMs) encode inputs using *times* that are dependent on the input itself. TEMs can therefore be used to model neurons or sensory receptors such as photoreceptors. In fact, the simpler neuron models often encode their input currents using action potentials with

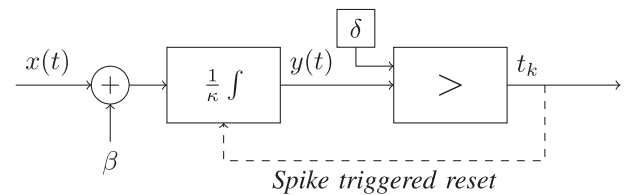


Fig. 2. Circuit of a Time Encoding Machine, with input $x(t)$, threshold δ , integrator constant κ and bias β .

fixed amplitude and varying timing, where the timing holds the information about the input [18].

In this paper, we will consider one model for time encoding machines which resembles an integrate-and-fire neuron with no leak [13], [18]. Such TEMs can provide perfect encodings of signals using one or many channels. The circuit of a TEM is depicted in Fig. 2.

Definition 1: A *time encoding machine* (TEM) with parameters κ , δ , and β takes an input signal $x(t)$, adds a bias β to it and integrates the result, scaled by $1/\kappa$, until a threshold δ is reached. Once this threshold is reached, the time t_ℓ at which it is reached is recorded, the value of the integrator resets to $-\delta$ and the mechanism restarts. We say that the machine spikes at the integrator reset and call the recorded time t_ℓ a *spike time*.

Note that a TEM can also be assumed to reset to zero after each spike with only minor effects on the results.

The first results on time encoding machines that resemble this model were, to the authors' knowledge, established by Lazar and Tóth [12].

The results operate under the following assumptions.

- A1) The input signal $x(t)$ is bandlimited with bandwidth Ω .
- A2) The input signal $x(t)$ is in $\mathcal{L}^2(\mathbb{R})$.

A3) The input signal $x(t)$ is bounded by a constant c , $|x(t)| < c, \forall t \in \mathbb{R}$.

Under these assumptions, the input $x(t)$ can be reconstructed from the emitted spike times if the parameters of the machine satisfy $\beta > c$ and the bandwidth satisfies

$$\Omega < \frac{\pi(\beta - c)}{2\kappa\delta}. \quad (1)$$

The reconstruction scheme and the proof of convergence are based on two key elements:

- 1) the time encoding scheme is tightly related to the scheme of sampling averages, therefore the results developed for the reconstruction from averages can be used for time encoding and reconstruction [14], [19], and
- 2) when performing time encoding, the maximal delay between two consecutive spike times is dictated by the parameters of the machine:

$$t_{\ell+1} - t_{\ell} < \frac{2\kappa\delta}{\beta - c}. \quad (2)$$

Given these two observations, and under Condition (1), the input signal can be perfectly determined by the spike times using algorithms based on alternating projections onto convex sets [9], [12], [20].

Following these findings, studies were conducted to evaluate the performance of different reconstruction approaches that use results from uniform sampling [21]–[23], that use iterative reconstruction which is implementable in hardware [9], [24], that run in an online fashion [25], [26] or that apply to more general classes of signals [7].

Later, the theory was extended to multi-channel time encoding of a signal. On one hand, Lazar suggested a scheme for bandlimited signal sampling and reconstruction using many time encoding machines coupled with filter banks [15]. On the other hand, we suggested a scheme for sampling and reconstructing a bandlimited signal using many time encoding machines that are similar and have no pre-filters [27]. In the latter scenario, we showed that if one TEM can encode a signal with bandwidth Ω , then M TEMs can encode a signal with bandwidth $M\Omega$ assuming that the TEMs have *unknown* non-zero shifts between their integrators $\alpha_1, \dots, \alpha_M$ [27]. In other words, M time encoding machines with parameters κ , δ , and β can encode a signal which satisfies assumptions (A1), (A2), (A3) if

$$\Omega < M \frac{\pi(\beta - c)}{2\kappa\delta}. \quad (3)$$

As a result, instead of using one TEM with a certain spiking rate to encode a signal, one can now use many TEMs with lower spiking rates to encode the same signal. This is useful if time encoding machines or neurons have an upper limit on their spiking rate.

The result generalizes to multi-signal, multi-channel time encoding, as partly studied in [28] and as we will see in the next sections.

This, in turn, allows us to understand how we can uniquely time encode video using multiple TEMs. Video time encoding machines have been examined before, and results on perfect reconstruction have already been established [29]. However, previous work relied on applying linearly independent filters

to the video before it is processed by the TEMs. As a result, one could deduce requirements on the minimal number of TEMs used. Here, we propose a filter-less approach where the scene is processed as is. Moreover, we clarify a dependency between spatial sampling density and temporal resolution that was not apparent before.

III. PROBLEM SETUP

For the remainder of this paper, we consider many time-varying signals $y^{(i)}(t), i = 1 \dots I$ that are correlated with each other. These signals are encoded using time encoding machines and we assume that each signal follows a parametric model which we know.

The goal is to recover the unknown inputs $y^{(i)}(t)$ from their time encoding.

Correlated signals arise in applications where principle component analysis yields little loss in information, such as, among others, meteorological data, biomarkers in human patients, regional economic data, audio, and video. The latter example will be given particular attention later on.

In our setup, we let $\mathbf{y}(t)$ denote the vector signal composed of $y^{(i)}(t)$'s and let $\mathbf{y}(t)$ be such that

- A4) each $y^{(i)}(t)$ has a finite parametric representation:

$$y^{(i)}(t) = \sum_{k=1}^K c_{i,k}(\mathbf{y}) f_k(t), \quad (4)$$

where the $c_{i,k}(\mathbf{y})$ are fixed coefficients that are unknown a priori and form a matrix $\mathbf{C}(\mathbf{y})$, and the $f_k(t)$'s, $k = 1 \dots K$ are known functions,

- A5) each $y^{(i)}(t)$ can be written as a linear combination of $x^{(j)}(t)$'s, $j = 1 \dots J$ where $J < I$:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t), \quad (5)$$

for $x^{(j)}(t)$'s characterized by a matrix of coefficients $\mathbf{C}(\mathbf{x})$ (where the $x^{(j)}(t)$'s and $\mathbf{C}(\mathbf{x})$ are unknown a priori) and a mixing matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$, and

- A6) each $y^{(i)}(t)$ is sampled using a time encoding machine TEM⁽ⁱ⁾ with parameters $\kappa^{(i)}$, $\delta^{(i)}$ and $\beta^{(i)}$ which are known and can vary between machines. The outputs of the machines are denoted $\{t_{\ell}^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}$.

The sampling setup we described is depicted in Fig. 3.

We will consider two options for the functions $f_k(t)$:

- (A7.a) $f_k(t)$ is a sinc function

$$f_k(t) = \text{sinc}_{\Omega}(t - \tau_k) = \frac{\sin(\Omega(t - \tau_k))}{\pi(t - \tau_k)}, \quad (6)$$

for Ω and τ_k known, so that the $y^{(i)}(t)$'s are a finite sum of sincs, or

- (A7.b) $f_k(t)$ is a complex exponential function,

$$f_k(t) = \exp\left(\mathbf{j} \frac{2\pi}{T} kt\right), \quad (7)$$

so that the $y^{(i)}(t)$'s are bandlimited periodic functions.

Thankfully, the functions in (A7.a) and (A7.b) resemble each other enough for the treatment of the two functions to be done

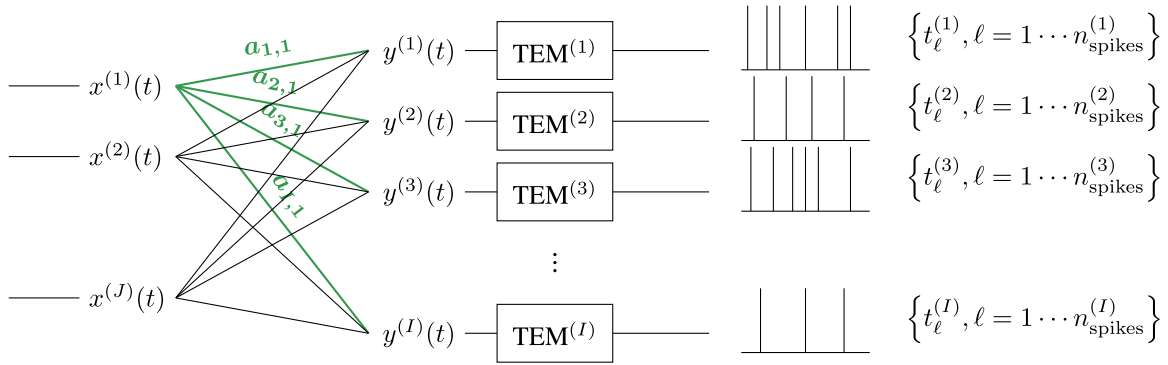


Fig. 3. Sampling setup: J input signals $x^{(j)}(t), j = 1 \dots J$ are mixed using a matrix \mathbf{A} and produce signals $y^{(i)}(t), i = 1 \dots I$. Each $y^{(i)}(t)$ is then sampled using a time encoding machine $\text{TEM}^{(i)}$ which produces spike times $\{t_\ell^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}$.

at the same time. For both of them, we consider the reconstruction conditions with \mathbf{A} satisfying either of the following two assumptions.

- (A8.a) The linear map from the low dimensional space $\mathbf{A} \in \mathbb{R}^{I \times J}$ is known.
- (A8.b) The linear map from the low dimensional space $\mathbf{A} \in \mathbb{R}^{I \times J}$ is unknown but the dimension of the low dimensional space J is known.

We first consider the case where \mathbf{A} is known and provide conditions for perfect reconstruction in Section IV-A and a reconstruction algorithm in Sections IV-B. We later provide applications for this scenario in Sections V and VI, where we deal with time encoding video.

Later, we will consider the case where \mathbf{A} is unknown and provide a reconstruction algorithm based on singular value projection for low-rank matrix recovery in Section VII. We then follow with simulations to show results and with example applications for time encoding time-varying scenes.

IV. KNOWN LOW-RANK FACTORIZATION: TIME ENCODING AND RECONSTRUCTION

A. Conditions for Perfect Reconstruction

We can establish the following sufficient conditions to ensure that a series of inputs $y^{(i)}(t)$ drawn at random are reconstructible from their time encoding using machines $\text{TEM}^{(i)}$.

Theorem 1: Let I signals $y^{(i)}(t), i = 1 \dots I$ satisfy assumptions (A4), (A5) and (A6), and their functions $f_k(t)$ satisfy either of (A7.a) or (A7.b) with the corresponding coefficients $c_{j,k}(\mathbf{x})$ being drawn from a Lipschitz continuous probability distribution. Now assume $\mathbf{A} \in \mathbb{R}^{I \times J}$ as defined in (A5) is known and has every J rows linearly independent. Then the inputs $y^{(i)}(t), i = 1 \dots I$ are exactly determined by the spike times $\{t_\ell^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}, i = 1 \dots I$, with probability one, if:

$$\sum_{i=1}^I \min(n_{\text{spikes}}^{(i)} - 1, K) > JK. \quad (8)$$

We can prove the above theorem by writing it as a problem of rank one measurements, also called bi-linear measurements in [30]. The full proof is provided in Appendix A.

B. Reconstruction Algorithm

The spike time outputs of the machines $\{t_\ell^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}$ provide constraints on the integral of the input signals:

$$\int_{t_\ell^{(i)}}^{t_{\ell+1}^{(i)}} y^{(i)}(u) du = 2\kappa^{(i)}\delta^{(i)} - \beta^{(i)}(t_{\ell+1}^{(i)} - t_\ell^{(i)}) =: \dot{b}_\ell^{(i)}. \quad (9)$$

These measurements can be rewritten to fit the rank one measurements formulation [30]. Letting $\mathbf{C}(\mathbf{x})$ denote the matrix of coefficients $c_{j,k}(\mathbf{x})$ for the underlying signals $x^{(j)}(t)$, we can reconstruct $\mathbf{C}(\mathbf{x})$ (and therefore $\mathbf{y}(t)$) by solving

$$\dot{b}_\ell^{(i)} = \text{vec} \left(\mathbf{a}_i \left[\dot{\mathbf{F}}_\ell^{(i)} \right]^T \right)^T \text{vec}(\mathbf{C}(\mathbf{x})), \quad (10)$$

where $\dot{b}_\ell^{(i)}$ is known, $\text{vec}()$ denotes the vectorization operation,

$$\left[\dot{\mathbf{F}}_\ell^{(i)} \right]_k = \int_{t_\ell^{(i)}}^{t_{\ell+1}^{(i)}} f_k(u) du,$$

and $t_1^{(i)}$ denotes the first spike time of $\text{TEM}^{(i)}$.

Under the conditions of Theorem 1, the linear system in (10) is full rank and $\mathbf{C}(\mathbf{x})$ can be recovered perfectly. Once the matrix $\mathbf{C}(\mathbf{x})$ has been recovered, one can recover the coefficients $c_{i,k}(\mathbf{y})$ of the $y^{(i)}(t)$'s by setting $\mathbf{C}(\mathbf{y}) = \mathbf{A}\mathbf{C}(\mathbf{x})$ and can therefore recover the original sampled signals.

Note that, given this formulation, there are different approaches to solving for $\mathbf{C}(\mathbf{x})$. One can either invert the system in (10) by using a pseudo-inverse, or use iterative approaches to minimize the mean-squared error, such as gradient descent or projections onto convex sets (POCS) [9], [14], [20] as we did in [28]. All of these methods converge to the same result, given that a pseudo-inverse approach minimizes the mean-squared error by definition, that the gradient descent approach is given a convex loss function and that the POCS approach uses convex sets which have a unique intersection. For the remainder of the manuscript, we use the pseudo-inverse approach, mostly for its speed of execution.

C. Interpretation

The result in Theorem 1 establishes a Nyquist-like criterion for recovery. It specifies how to count the number of linearly independent constraints in the multi-channel TEM setup and requires as many of these constraints as there are degrees of freedom to recover the sampled signals. Given that each *pair* of spike times corresponds to one linear constraint on the input, the results can be summarized by a few key points:

- 1) When sampling a collection of signals with a known linear mapping to or from a lower dimensional representation, what matters is the number of degrees of freedom in the low dimensional space, rather than the number of degrees of freedom in the high dimensional space. More practically, to ensure perfect reconstruction, we need the number of linearly independent constraints to be at least the number of degrees of freedom in the low dimensional space JK . In the case where $J \ll I$, we can see how this can be a major improvement in spiking rate.
- 2) When multiple correlated signals are sampled using different time encoding machines, a lower spiking rate of one machine can be compensated for by higher spiking rates from others. This can be seen by observing the summation in (8) and noting that the *total* spiking rate of the machines matters more than the individual spiking rates.
- 3) One machine can only compensate for another machine's low spiking rate up to a certain degree. This can be seen by the min term in (8) which implies that every machine has a maximal "useful" spiking rate depending on the signal and that going above this spiking rate does not add further information.

This has a series of implications. First, signals that have lower dimensional representations can be sampled at lower rates overall, increasing sampling efficiency. Second, if TEMs have limited capacity in terms of spiking rates (for example they have a refractory period), this can be compensated for by adding more TEMs. This would still ensure reconstruction of the input since the reconstruction condition in (8) is only linked to the number of degrees of freedom in the low dimensional space. Third, we will see in Section VI how the results help us solve time encoding of time-varying spatial signals which have certain structure in space.

Note that these results provide a stark improvement to sampling high-dimensional but low-complexity signals using regular clock-based sampling. In fact, Theorem 1 holds because of one key element: different dimensions of the signal are sampled at *different* times with continuous probability distributions. Uniform and synchronous sampling does not have this property; indeed, it only takes a short mental exercise to see that the recovery of $\mathbf{y}(t)$ takes IK samples if the $y^{(i)}(t)$'s are all sampled at the same sampling times.

To be fair, one *could* ensure that different $y^{(i)}(t)$'s are sampled at different times (minus the continuous probability condition), but this condition is much more elegantly ensured in the time encoding scenario. Moreover, using different clocks in the classical sampling setup poses difficulties because it is hard to align them. Clock alignment is not an issue in time encoding: the

outputs are trains of spikes and finding delays between TEMs is solved by adding spike trains and comparing spike times on one time axis.

D. Extensions

The results in this section assumed a fixed signal model and relied on two key elements: the structure of the matrix \mathbf{A} and the timing of the spikes which is asynchronous across machines. In this case, the asynchrony of the spikes across TEMs occurs because the matrix \mathbf{A} has different rows. However, this asynchrony can also be ensured by TEMs having different parameters, as done in [27]. In such a scenario, repeated (or linearly dependent) entries in \mathbf{A} are allowed and our assumptions can be relaxed, so that \mathbf{A} does not have to have every J rows linearly independent. We choose not to further explore this possibility, here, but rather to provide a basic framework to understand time encoding of mixed signals.

V. REPRESENTING 2D SIGNALS WITH SPIKES: VIDEOS WITH 1 SPATIAL DIMENSION

The results obtained in previous sections provide considerable improvements in sample requirements for multi-signal reconstruction when these signals have a low dimensional structure.

Interestingly, the problem of time encoding bandlimited video can be rephrased to fit this framework.

First, we start with a simpler case and consider a two-dimensional (2D) signal $y(d, t)$ that is bandlimited in both components. Note that, exclusively when using the terms "2D" and "3D," when we refer to "dimension," we mean the spatial and time dimensions, i.e. the signal varies along each of these two components. We do not refer to the complexity of the signal (as it relates to I , J and K) as we did before.

$$\begin{aligned} y(d, t) &= \sum_{k_0=-K_0}^{K_0} \sum_{k_1=-K_1}^{K_1} c_{k_0, k_1}(y) \exp\left(j2\pi\left(\frac{tk_0}{T} + \frac{dk_1}{D}\right)\right) \\ &= \sum_{k_0=-K_0}^{K_0} \sum_{k_1=-K_1}^{K_1} c_{k_0, k_1}(y) \\ &\quad \exp\left(j2\pi\left(\frac{tk_0}{T}\right)\right) \exp\left(j2\pi\left(\frac{dk_1}{D}\right)\right), \end{aligned}$$

where the $c_{k_0, k_1}(y)$'s denote the 2D Fourier series coefficients of $y(d, t)$. Note that we assume that $y(d, t)$ has $(2K_0 + 1) \times (2K_1 + 1)$ of these coefficients with periods T and D in the time and space components, respectively.

The results here will concern any such signal but, to make the treatment more intuitive, we will assume that we are dealing with a visual scene that has one continuous spatial component d and is varying along time t . To be clearer, taking a picture of this scene at time t provides the light intensity along one direction, which we assume to be the horizontal direction, without loss of generality. See Fig. 4 for illustration.

Now assume that we sample this time-and-horizontally-varying scene using I TEMs. Each TEM^(*i*) is associated with a

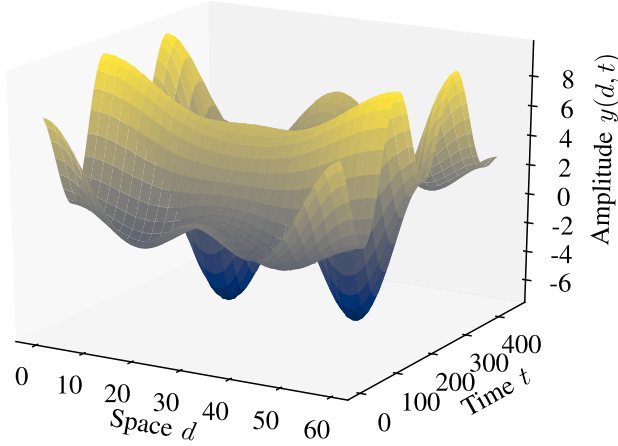


Fig. 4. 2D signal $y(d, t)$ with spatial component d and time component t .

location in “space,” i.e. a position on the horizontal axis, $d^{(i)}$, such that the sampled signal $y^{(i)}(t)$ satisfies:

$$y^{(i)}(t) = y(d^{(i)}, t). \quad (11)$$

To make the connection to the theory in Section IV, we first need to define an auxiliary vector signal $\mathbf{x}(t)$ with $2K_1 + 1$ components, such that

$$x^{(k_1)}(t) = \sum_{k_0=-K_0}^{K_0} c_{k_0, k_1}(y) \exp\left(j \frac{2\pi}{T} k_0 t\right). \quad (12)$$

We now notice that we can rewrite

$$y^{(i)}(t) = y(d^{(i)}, t) = \sum_{k_1=-K_1}^{K_1} x^{(k_1)}(t) \exp\left(j \frac{2\pi}{D} k_1 d^{(i)}\right). \quad (13)$$

We can directly see that this brings back the structure we saw earlier, we have $\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t)$ where $\mathbf{x}(t)$ is as defined in (12) and $[\mathbf{A}]_{i, k_1} = \exp(-j \frac{2\pi}{D} k_1 d^{(i)})$, where i denotes the sampled channel.

VI. REPRESENTING 3D SIGNALS WITH SPIKES: VIDEOS WITH 2 SPATIAL DIMENSIONS

A. Theory

We can use a similar treatment to understand how to time encode and reconstruct 3D signals $y(d_1, d_2, t)$. These signals can be interpreted as scenes that have 2 spatial components d_1 and d_2 (horizontal and vertical) and one time component t , as in videos.

We again assume that such a signal $y(d_1, d_2, t)$ is bandlimited along all components:

$$y(d_1, d_2, t) = \sum_{k_0=-K_0}^{K_0} \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} c_{k_0, k_1, k_2}(y) \exp\left(j 2\pi \left(\frac{t k_0}{T} + \frac{d_1 k_1}{D_1} + \frac{d_2 k_2}{D_2}\right)\right) \quad (14)$$

Once again we assume that $y(d_1, d_2, t)$ is sampled in space at locations specified by \mathbf{d} where sample i is taken at spatial

location $\mathbf{d}^{(i)} = (d_1^{(i)}, d_2^{(i)})$ for some $d_1^{(i)}$ and $d_2^{(i)}$ in \mathbb{R} . An example is provided in Fig. 5.

We define $x^{(k_1, k_2)}(t)$ in a similar fashion to (12):

$$x^{(k_1, k_2)}(t) = \sum_{k_0=-K_0}^{K_0} c_{k_0, k_1, k_2}(y) \exp\left(j 2\pi \left(\frac{t k_0}{T}\right)\right), \quad (15)$$

and we obtain the input signals to the TEMs

$$y^{(i)}(t) = \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} x^{(k_1, k_2)}(t) \exp\left(j 2\pi \left(\frac{d_1^{(i)} k_1}{D_1} + \frac{d_2^{(i)} k_2}{D_2}\right)\right). \quad (16)$$

Once more, we have found that we are time encoding $\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t)$ with the entries of $\mathbf{x}(t)$ satisfying (15), and a matrix \mathbf{A} which is known if we know the locations of the time encoding machines $\mathbf{d}^{(i)}$. If the locations of the time encoding machines $\mathbf{d}^{(i)}$ are such that \mathbf{A} has every $(2K_1 + 1)(2K_2 + 1)$ rows linearly independent, then all coefficients $c_{k_0, k_1, k_2}(t)$ can be recovered using $\prod_{n=0}^2 (2K_n + 1)$ appropriate measurements. This means that the continuous scene can also be recovered, so we can interpolate the scene between spike times in both space and time components.

One example of a matrix \mathbf{A} that satisfies the above constraint arises when one follows *sufficient uniform gridding*.

Definition 2: Sufficient Uniform Gridding defines the sampling locations $d^{(i)}$ to follow a uniform grid over a spatial period, with $2K_1 + 1$ positions in the d_1 direction and $2K_2 + 1$ positions in the d_2 direction. More formally, i ranges between zero and $(2K_1 + 1)(2K_2 + 1)$ and

$$d_1^{(i)} = \frac{\lfloor i / (2K_2 + 1) \rfloor}{2K_1 + 1} D_1, \quad (17)$$

$$d_2^{(i)} = \frac{(i \bmod (2K_2 + 1))}{2K_2 + 1} D_2. \quad (18)$$

Lemma 1: The matrix \mathbf{A} obtained from using sufficient uniform gridding with entries as defined in (16) has every $(2K_1 + 1)(2K_2 + 1)$ rows linearly independent.

Proof: The proof of Lemma 1 relies on calculating the Gram matrix of \mathbf{A} and noticing that it is diagonal and therefore full rank. Consequently \mathbf{A} also has a full rank $(2K_1 + 1)(2K_2 + 1)$ and has every $(2K_1 + 1)(2K_2 + 1)$ rows linearly independent. ■

This is not the only case in which \mathbf{A} satisfies our assumptions, it seems that more general configurations of the spatial sampling can also work provided the samples cover the space.

As was the case in Section IV, admitting that TEMs are receiving input signals that have a low dimensional structure allows one to manipulate the number of time encoding machines while keeping the same total spiking rate, and without compromising on reconstructibility.

In other words, every TEM does not have to be able to perfectly reconstruct its own input for the entire scene to be reconstructed. On the contrary, emitted spikes from all machines

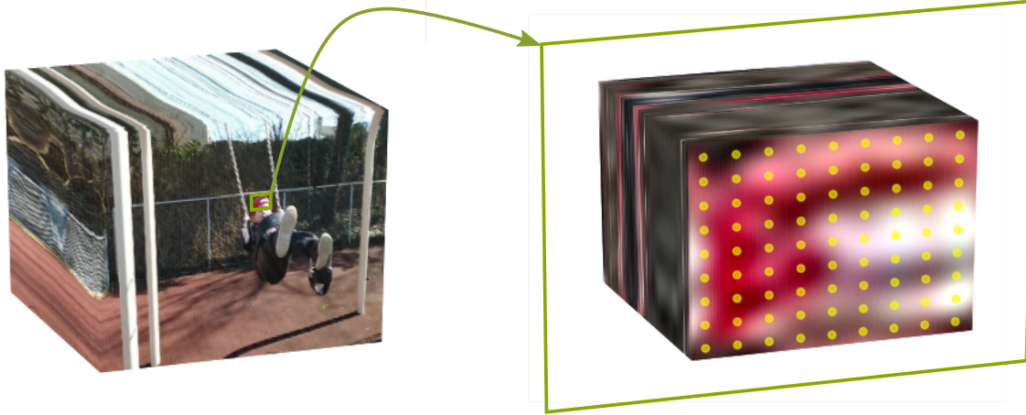


Fig. 5. A time-varying scene is sampled at different spatial locations using TEMs. On the left, we see the scene with varying spatial and time components taken from the Need for Speed dataset [8]. On the right, we see a time-varying patch which we will record using time encoding machines placed at the yellow dots. Originally, the video data we use was captured using a standard frame-based camera. We smoothly interpolate the video by assuming that the underlying structure is bandlimited and periodic and we aim to estimate the corresponding Fourier Series coefficients using the spikes emitted by the TEMs. In this case, we use a 9×9 grid of TEMs on an interpolated version of a $9 \times 9 \times N_f$ video patch where N_f is the number of frames used for the interpolation and time encoding. Therefore the number of Fourier Series coefficients to obtain is $9 \times 9 \times N_f$.

are used collaboratively in order to reconstruct the scene which has a parametric representation.

Therefore, if we have TEM-like receptors or sensors that have a limited spiking rate, spatial and temporal resolution can be regained by adding more sensors at new locations.

B. Simulations

We would like to illustrate the theoretical results obtained in the previous section on an actual video, to illustrate the relationship between spatial and temporal sampling density. First we choose a video recorded with a standard frame-based camera [8], and examine a patch of this video as shown in Fig. 5. This patch has $H \times W \times N_f$ samples where H refers to the height of the patch in pixels, W refers to the width of the patch in pixels and N_f refers to the number of frames.

We assume the underlying scene has a periodic bandlimited structure (which is also the assumption that allows for finite uniform sampling). This allows us to (1) express the scene as in 14 and (2) fix the corresponding number of Fourier series coefficients to match the number of samples $(2K_0 + 1) \times (2K_1 + 1) \times (2K_2 + 1)$ where $K_0 = \lfloor H/2 \rfloor$ for example. The patch we consider is therefore a smooth function with a fixed number of parameters we are interested in and which can be sampled anywhere in time and space.

Given the smoothly varying patch, we place TEMs, for example, at the yellow dots in Fig. 5. In this case, we have a patch which is 9 pixels high and 9 pixels wide and we place a 9×9 grid of time encoding machines, according to the definition of sufficient uniform gridding. We will show, in our experiments that this is the minimum number of TEMs required to achieve perfect reconstruction.

We will also show how we can use more TEMs in the spatial components to obtain better resolution in the time component. This will not necessarily be the case the other way around: more sampling in time does not always provide improved spatial frequency resolution.

The interpolated patch from Fig. 5 is sampled using a fixed number of TEMs and we vary the number of spikes per TEM to see how this effects the reconstruction error in the left part of Fig. 6.

We consider three scenarios from top to bottom: we have a 9×15 uniformly spaced grid of TEMs, a 9×9 uniformly spaced grid of TEMs (similar to uniform sufficient gridding), and a 9×5 uniformly spaced grid of TEMs.

We examine the evolution of the reconstruction error as the number of spikes per TEM increases. Having JK constraints provided by the spikes (dashed green lines in the figure) will not always ensure that the reconstruction error decrease significantly. The results rather match the predictions of Theorem 1, as indicated by the vertical orange line. Moreover, as validated by our experiments, Theorem 1 cannot place any guarantees on reconstruction for the case where there are fewer TEMs than spatial components (i.e. when we have a 9×5 grid of TEMs). In fact, perfect reconstruction is never possible: the system will always be underdetermined because of too few sensors in the spatial domain.

On the other hand, we examine the scenario where we vary the number of TEMs for a fixed spiking rate per machine in the right part of Fig. 6. We similarly set the spiking rate to different levels from top to bottom: 5 spikes per TEM, 9 spikes per TEM and 15 spikes per TEM. Here, the sufficient number of spikes per machine $2K_0 + 1 = 9$ is the one that allows each machine to perfectly resolve its own input.

We notice that the reconstruction error undergoes a significant decrease once the number of TEMs is such that the condition of Theorem 1 is satisfied. As was the case for the left part of Fig. 6, the threshold at which this decrease occurs does not depend on the total number of constraints (in green) but rather on the number of *linearly independent* constraints.

We can draw a similar conclusion to that drawn for the left part of Fig. 6: increasing the number of spikes per TEM beyond a certain point is not helpful and it is generally more beneficial to have more TEMs or sensors that spike less frequently.

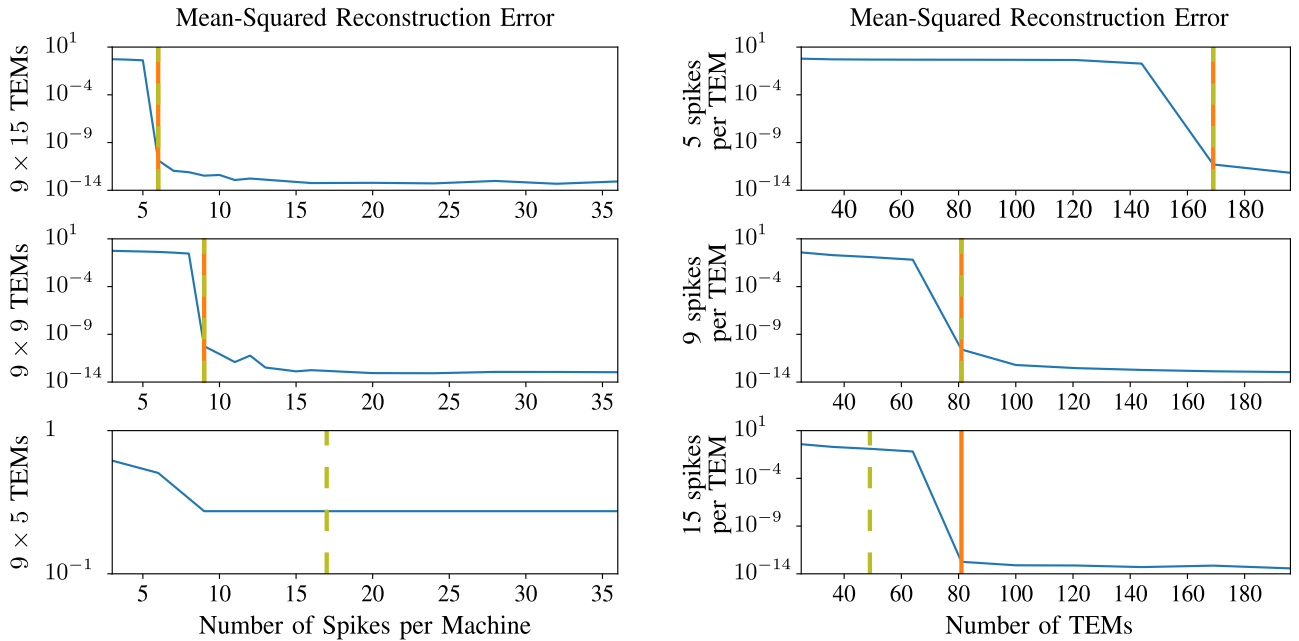


Fig. 6. On the left, we show the mean-squared reconstruction error as the number of spikes per time encoding machine varies. On the right, we show the mean-squared reconstruction error as the number of time encoding machine varies for fixed numbers of spikes. We assume the video has $9 \times 9 \times 9$ Fourier series coefficients that we wish to recover. On the left-hand side, we show, from top to bottom, the evolution of the error as number of spikes increases for 9×15 , 9×9 and 9×5 uniformly spaced TEMs. Note that the second row assumes sufficient uniform gridding. On the right-hand side, we show, from top to bottom, the evolution of the error as number of spikes increases for 5, 9 and 15 spikes emitted per machine. Note that the second row assumes a spiking rate which matches the sufficient rate starting from which each TEM can perfectly reconstruct its input. For each plot, the dashed green line marks the number of spikes per machine or the number of TEMs (for the left and right hand sides respectively) starting from which we have more constraints than unknowns, not accounting for linear independence. The vertical orange line instead marks the threshold provided by Theorem 1 and marks the number of spikes per TEM (left) or the number of TEMs (right) starting from which we have more *linearly independent* constraints than unknowns. Note that, if we assume no underlying structure, one would always need at least 9 TEMs in each spatial direction and at least 9 spikes per time encoding machine.

C. Coupling of Spatial and Temporal Resolution: Intuition and Consequences

In a nutshell, the theory developed and experiments conducted all indicate that, if one would like to increase resolution, whether spatial or temporal, it is better to increase spatial sampling density. Increasing spatial sampling density is always useful, unlike increasing the number of spikes per machine, as indicated by the condition in (8) of Theorem 1.

In fact, a TEM can only output as much information as it receives, so if a TEM perfectly characterizes its own input using 15 spikes, there is no point in generating 20, 30 or 40 spikes.

On the other hand, increased spatial sampling can aid spatial *and* temporal resolution because TEMs located at different locations will almost surely spike at different *times* because they either have different inputs or different initial conditions [27], or both.

This particular characteristic is not met by standard, frame-based video recordings where all pixels record information at the same time. Unfortunately, when all information is recorded at the same time, any information obtained from oversampling in the spatial domain is redundant and every pixel should be able to reconstruct its input on its own. On the other hand, with firing times that are asynchronous across sensors, more spatial information contributes to better resolution in the time domain and pixels can collaborate to reconstruct the scene.

In practice, this means two things: (1) TEMs or event-based sensors that have a limited spiking rate can be compensated for by simply having more sensors in space and (2) it is better to increase sampling capacity in the spatial domain when performing time encoding because this can improve *both* spatial and temporal resolution.

Of course, one can theoretically also achieve the same result with pixels that sample uniformly at different times. However, this can pose problems in practice because it can be difficult to determine the shift of out-of-sync uniform samplers when they do not have the same clock. This is not a problem with time encoding because TEMs output events which hold information in their timing. The clocks of TEMs can therefore be aligned by summing the outputs such that events are timed using the same clock [27]. That being said, if the shift between uniform samplers can be perfectly determined, one would obtain similar effects as exposed here. Our main message is that the frame-based video paradigm that resulted from stacking images can be improved by using out-of-sync samplers. TEMs naturally provide out-of-sync sampling in addition to providing other advantages such as low power consumption and an input-dependent spiking rate.

VII. UNKNOWN LOW-RANK FACTORIZATION

A. Problem Formulation and Algorithm

We revisit the setup exposed in Section III. So far, we have assumed that we are given the time encodings of a collection of

Algorithm 1: Singular Value Projection.

Input: \mathcal{S} , b , tolerance ϵ , η_m for $m = 0, 1, 2, \dots$

- 1: $X^0 = 0$ and $m = 0$
- 2: **repeat**
- 3: $Y^{m+1} \leftarrow X^m - \eta_m \mathcal{S}^T(\mathcal{S}(X^m) - b)$
- 4: Compute top J singular vectors of Y^{m+1} : U_J, Σ_J, V_J
- 5: $X^{m+1} \leftarrow U_J \Sigma_J V_J^T$
- 6: $m \leftarrow m + 1$
- 7: **until** $\|\mathcal{S}(X^{m+1}) - b\|_2^2 \leq \epsilon$

signals $y^{(i)}(t)$ with a low dimensional structure which we can reach by a known linear transformation $\mathbf{A} \in \mathbb{R}^{I \times J}$ and that we are asked to reconstruct the inputs $y^{(i)}(t)$. While this is a useful model in itself, we are also interested in studying the case where the linear transform \mathbf{A} is unknown.

Once again, we assume we have the time encodings of a collection of signals $y^{(i)}(t)$ which satisfy assumptions (A4), (A5) and (A6). Furthermore, we assume the functions $f_k(t)$ of $y^{(i)}(t)$ satisfy either of (A7.a) or (A7.b) and that the linear transformation \mathbf{A} is unknown as in (A8.b).

We wish to recover the signals $y^{(i)}(t)$, $i = 1 \dots I$ from their time encoding, with as few samples as possible.

To do so, we aim to reconstruct the coefficients of the parametric representation of $\mathbf{y}(t)$, $c_{i,k}(\mathbf{y})$ as defined in (A4). These coefficients are placed in the matrix $\mathbf{C}(\mathbf{y})$, with row i containing the coefficients of signal $y^{(i)}(t)$. We note once more that $\mathbf{C}(\mathbf{y})$ can be written:

$$\mathbf{C}(\mathbf{y}) = \mathbf{A}\mathbf{C}(\mathbf{x})$$

where $\mathbf{A} \in \mathbb{R}^{I \times J}$, $\mathbf{C}(\mathbf{x}) \in \mathbb{R}^{J \times K}$, $J < I$ and J is known.

In words, $\mathbf{C}(\mathbf{y})$ is a matrix which has a low rank matrix decomposition with a known rank.

The matrix $\mathbf{C}(\mathbf{y})$ is probed using a sensing operator which we will call \mathcal{S} . The sensing operator performs the measurements in (10), i.e.

$$\mathcal{S}_n(\mathbf{C}(\mathbf{y})) = \dot{b}_n, \quad (19)$$

where we index a pair (i, ℓ) by n so that $\dot{b}_n = \dot{b}_\ell^{(i)}$ is as defined in (9)

Given this measurement setup, we can adopt the Singular Value Projection approach to recover the matrix $\mathbf{C}(\mathbf{y})$ from few measurements [31].

The Singular Value Projection (SVP) algorithm alternately applies the low-rank constraint and the measurement constraint on the matrix of interest $\mathbf{C}(\mathbf{y})$. In Algorithm 1 we let X^m be the estimate at iteration m of the target matrix to reconstruct (in our case this is $\mathbf{C}(\mathbf{y})$) and Y^m be a proxy matrix to perform the iterations.

The SVP algorithm is based on projected gradient descent. Reconstruction guarantees for this algorithm were initially established in cases where the sensing operator satisfies the Restricted Isometry Property [31]–[33]. This property does not hold in our case, given that our measurement operators \mathcal{S}_n have rank one. The rank one scenario has been treated in [34] where

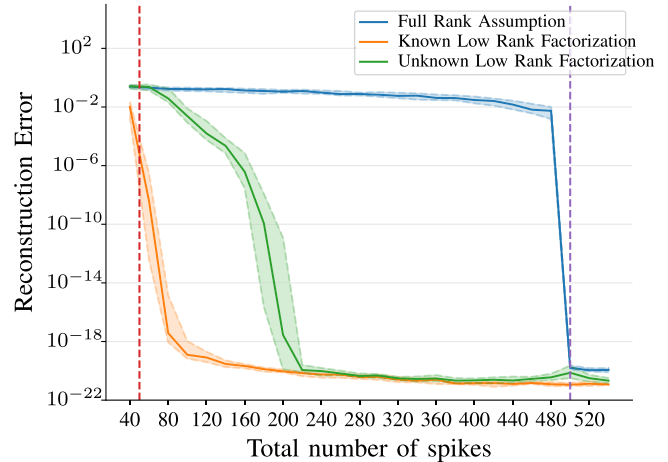


Fig. 7. Reconstruction error of one out of twenty signals that have rank two, as the number of emitted spikes increases. The red dashed line marks the perfect reconstruction condition assuming the transform to the low dimensional space is known, and the purple dashed line marks the perfect reconstruction condition assuming there is no lower dimensional representation of the signals. We show the median and quartiles of the reconstruction error for 25 random trials, when assuming the signals have no low dimensional structure, when assuming they have a low dimensional structure with a known linear mapping, and when they have a low dimensional structure with an unknown linear mapping.

Gaussianity assumptions are made. Again, these assumptions do not hold for our case and we leave the theoretical analysis of convergence for future work. We do, however, illustrate the utility of our approach with simulations in the next section.

B. Simulations

We provide simulation results to evaluate the reconstruction performance in different regimes. We consider the scenario where we time encode and reconstruct twenty signals that are composed of 25 sinc functions at known locations and that can be written as linear combinations of two such signals.

We evaluate the reconstruction performance that varies as the number of spikes of all machines increase uniformly. We do this in the following cases:

- (S1) when assuming the signals have no underlying low dimensional structure
- (S2) when assuming the signals have an underlying low dimensional representation which we can reach through a known linear transform \mathbf{A} , and
- (S3) when the signals have an underlying low dimensional representation with an unknown mapping \mathbf{A} .

For each of these cases, we draw the entries of $\mathbf{C}(\mathbf{x})$ and \mathbf{A} uniformly at random and time encode and reconstruct all twenty signals, computing the obtained normalized mean-squared error for the first signal among the twenty, assuming a random mapping \mathbf{A} to low dimensional space. Then we plot the median and quartiles of the mean-squared error on a log plot to compare performance. Results are included in Fig. 7.

Note that, if we assume no underlying low dimensional structure (S1), the signals should be reconstructible assuming there are $I \times K$ linearly independent constraints. In this case, since

the number of spikes of all machines increase uniformly, we will need $I \times K = 500$ spikes. As for the scenario (S2), according to Theorem 1, the signals should be reconstructible assuming that there are $J \times K$ linearly independent constraints. As before, this means we would need $J \times K = 50$ spikes.

We draw each of these conditions in Fig. 7 to see if the performance is consistent with our expectations.

Assuming we know the transformation \mathbf{A} to a low-dimensional space (S2) greatly improves reconstruction compared to when we assume that there is no low rank structure for the input (S1): the error decays much earlier in the first case than it does in the second case.

Assuming such a transformation exists but that we do not know it (S3), also offers benefit. While the reconstruction algorithm can be quite unstable in regimes where the number of spikes is not sufficient, it can yield a very good reconstruction for a higher number of spikes, where the scenario (S1) fails entirely.

VIII. CONCLUSION

We have shown how time encoding can be used to encode and reconstruct multiple signals that have lower-dimensional representations.

The general case can be treated by reformulating our problem as a rank-one matrix measurement problem: we have shown that signals that have a known lower dimensional representation require fewer spikes for perfect reconstruction than if this lower dimensional representation did not exist.

Time encoding videos can then be rewritten as a special case of low-rank signal estimation. As a consequence, we show through theory and experiments that, if one wishes to increase spatial or temporal resolution, it is better to sample densely in space than to have TEMs emit more spikes. More practically, in the case of an event-based camera, it is better to have more pixels that fire asynchronously than to have pixels that fire more often.

Finally we have also examined the case where the signals of interest are low rank but we do not know the transformation to the low rank space. We applied low rank factorization algorithms and found significant experimental improvements compared to the case where no low rank structure is assumed.

In future work, we would like to further investigate low rank factorization within the time encoding setup and understand how it can be used to encode multi-dimensional data with a different structure to that presented in the paper.

APPENDIX A

KNOWN LOW DIMENSIONAL MAPPING - ELABORATION AND PROOF OF THEOREM 1

To prove Theorem 1, we will use results about rank-one matrix measurements [30]. The work in [30] assumes that one is attempting to reconstruct a matrix \mathbf{C} using measurements of the form:

$$b_n = \mathbf{g}_n^T \mathbf{C} \mathbf{h}_n, \quad (20)$$

and rewrites the measurements as

$$b_n = \text{vec}(\mathbf{g}_n \mathbf{h}_n^T)^T \text{vec}(\mathbf{C}), \quad (21)$$

where $\text{vec}(\cdot)$ is the vectorization operator.

Note that we adopted a change of notation with respect to [30] to avoid confusion.

The results of [30] then hold under two further assumptions.

A9) \mathbf{h}_n can be parametrized by one variable $t \in \mathbb{R}$. More precisely, we assume the k -th entry of h_n has the form $[h_n]_k = h_k(t_n)$ where $h_k : \mathcal{I} \rightarrow \mathbb{R}, k = 0, \dots, K-1$ are linearly independent functions from a linear space of functions \mathcal{F} , $\mathcal{I} \in \mathbb{R}$ is an interval or the whole real line and $t_n \in \mathcal{I}, n = 0, \dots, N-1$ are sampling times. Moreover, it is assumed that the sampling times (t_0, \dots, t_{N-1}) follow a continuous probability distribution on \mathcal{I}^N and that for every non-zero element $h \in \mathcal{F}$, the set of zeros of h has Lebesgue measure (λ) equal to zero: $\lambda(\{t | f(t) = 0\}) = 0$.

A10) The vectors \mathbf{g}_n are taken from a set \mathcal{A} , where every J elements of \mathcal{A} are linearly independent.

As a result, a uniqueness condition can be obtained.

Theorem 2 (Pacholska '20): Consider the set of KJ vectors of the form $\text{vec}(g_n h_n^T)$. It is a basis in \mathbb{R}^{KJ} if and only if no more than K vectors g_n are equal.

We are able to rewrite our problem as a rank-one measurement problem.

We will use the following lemmas to prove Theorem 1.

Lemma 2: Under the assumptions of Theorem 1, the spike times $\{t_\ell^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}, i = 1 \dots I$ follow a continuous probability distribution.

Proof: This closely follows the proof in [30]. \blacksquare

Lemma 3: Let the $f_k(t)$'s be the functionals as defined in (A7.a) or (A7.b) and define $F_k(t) = \int_{t_0}^t f_k(u) du$. Then the F_k 's are linearly independent functions from a linear space of functions \mathcal{F} which is the space of bandlimited functions. Moreover, every non-zero element of \mathcal{F} has a set of zeros with Lebesgue measure equal to zero.

Proof: This follows by construction of the f_k 's which are linearly independent, leading to their integrals being linearly independent. The second part of the lemma follows from both sums of complex exponentials and sinc functions having a countable number of zeros. The sum of complex exponentials has a countable number of zeros because it can be rewritten as a polynomial, which has zeros described by the fundamental theorem of algebra, and the sum of sincs can also be written as a polynomial of a function of time, divided by the time, which also yields countable zeros [35]. Given that the set of zeros of each of these functions is countable, they have Lebesgue measure zero [36]. \blacksquare

Before proving Theorem 1, we prove the following lemma.

Lemma 4: Let I signals $y^{(i)}(t), i = 1 \dots I$ satisfy assumptions (A4), (A5) and (A6), and their function $f_k(t)$ satisfy either of (A7.a) or (A7.b) with the corresponding coefficients $c_{j,k}(\mathbf{x})$ being drawn from a Lipschitz continuous probability distribution. Now assume $\mathbf{A} \in \mathbb{R}^{I \times J}$ as defined in (A5) is known and has every J rows linearly independent. Then the inputs $y^{(i)}(t), i = 1 \dots I$ are exactly determined by the spike times

$\{t_\ell^{(i)}, \ell = 1 \dots n_{\text{spikes}}^{(i)}\}, i = 1 \dots I$ if:

$$\sum_{i=1}^I \min(n_{\text{spikes}}^{(i)}, K) > JK, \quad (22)$$

if the time encoding machines start sampling at t_0 with a known integrator value $\zeta_0^{(i)}$.

The integrator value $\zeta_0^{(i)}$ indicates the value of the integral of TEM⁽ⁱ⁾ at time t_0 , before the time encoding begins.

Proof of Lemma 4: We will assume that we operate under the assumptions set out in Lemma 1.

We start by showing that different constraints imposed by the time encoding machines can be written as in (20). In fact, two consecutive spike times $t_\ell^{(i)}$ and $t_{\ell+1}^{(i)}$ from a machine TEM⁽ⁱ⁾ impose a constraint on the integral of the concerned signal:

$$\int_{t_\ell^{(i)}}^{t_{\ell+1}^{(i)}} y^{(i)}(u) du = 2\kappa^{(i)}\delta^{(i)} - \beta^{(i)}(t_{\ell+1}^{(i)} - t_\ell^{(i)}) = b_\ell^{(i)}. \quad (23)$$

We define $Y^{(i)}(t) = \int_{t_0}^t y^{(i)}(u) du$ to be the integral of the signal $y^{(i)}(t)$ between t_0 and any later time t . Given (23), and that we know the initial integrator value $\zeta_0^{(i)}$, we can compute $Y^{(i)}(t_\ell^{(i)})$ for any spike time $t_\ell^{(i)}$:

$$Y^{(i)}(t_\ell^{(i)}) = \int_{t_0}^{t_\ell^{(i)}} y^{(i)}(u) du = 2\ell\kappa^{(i)}\delta^{(i)} - \beta^{(i)}(t_\ell^{(i)} - t_0) + \zeta_0^{(i)}. \quad (24)$$

We define this quantity to be $b_\ell^{(i)} := Y^{(i)}(t_\ell^{(i)})$ and denote the function $F_k(t) = \int_{t_0}^t f_k(u) du$. We then rewrite the right-hand side of (24) in terms of the parametrization of $y^{(i)}(u)$:

$$\begin{aligned} b_\ell^{(i)} &= \int_{t_0}^{t_\ell^{(i)}} \sum_{k=1}^K c_{i,k}(\mathbf{y}) f_k(u) du \\ &= \sum_{k=1}^K c_{i,k}(\mathbf{y}) \int_{t_0}^{t_\ell^{(i)}} f_k(u) du \\ &= \sum_{k=1}^K c_{i,k}(\mathbf{y}) F_k(t_\ell^{(i)}) \\ &= [\mathbf{F}(t_\ell^{(i)})]^T [\mathbf{C}(\mathbf{y})]_i^T \end{aligned} \quad (25)$$

Where we defined $[\mathbf{F}(t_\ell^{(i)})]$ to be the vector of integrals $F_k(t_\ell^{(i)})$ for $k = 1 \dots K$. We also defined $\mathbf{C}(\mathbf{y})$ to be the matrix of coefficients $c_{i,k}(\mathbf{y})$ as defined in (A4) and $[\mathbf{C}(\mathbf{y})]_i$ is the i^{th} row containing the coefficients for signal $y^{(i)}(t)$.

We further rewrite $\mathbf{C}(\mathbf{y}) = \mathbf{A}\mathbf{C}(\mathbf{x})$ (from (A5)) and obtain $[\mathbf{C}(\mathbf{y})]_i = [\mathbf{A}]_i \mathbf{C}(\mathbf{x})$. We thus obtain:

$$\begin{aligned} b_\ell^{(i)} &= [\mathbf{F}(t_\ell^{(i)})]^T ([\mathbf{A}]_i \mathbf{C}(\mathbf{x}))^T \\ b_\ell^{(i)} &= [\mathbf{F}(t_\ell^{(i)})]^T \mathbf{C}(\mathbf{x})^T [\mathbf{A}]_i^T \end{aligned}$$

$$b_\ell^{(i)} = [\mathbf{A}]_i \mathbf{C}(\mathbf{x}) [\mathbf{F}(t_\ell^{(i)})] \quad (26)$$

We can thus reindex the above equations: we let every n correspond to a single pair (ℓ, i) and let $b_n = b_\ell^{(i)}$, $\mathbf{g}_n = [\mathbf{A}]_i$ and $\mathbf{h}_n = [\mathbf{F}(t_\ell^{(i)})]$.

We can now see that the vectors \mathbf{h}_n can be parametrized by one variable $t \in \mathbb{R}$ using a set of functions h_k which satisfy assumption (A9), as stated by Lemma 3. Moreover, according to Lemma 2, the spike times follow a continuous probability distribution, as required in assumption (A9).

We can also see that the vectors \mathbf{g}_n just defined satisfy assumption (A10) by construction since this is a condition in Lemma 4.

Then, we note that under the conditions of Lemma 4, one can extract JK constraints that satisfy the constraints of Theorem 2, thus ensuring perfect reconstruction of the matrix of parameters $\mathbf{C}(\mathbf{x})$. ■

Proof of Theorem 1: Using similar notation used for the Proof of Lemma 4, we note that the value $b_\ell^{(i)}$ is not known, when the initial integrator values $\zeta_0^{(i)}$ are not known, instead, we know the value of

$$\tilde{b}_\ell^{(i)} = b_\ell^{(i)} + \zeta_0^{(i)} = [\mathbf{A}]_i \mathbf{C}(\mathbf{x}) [\mathbf{F}(t_\ell^{(i)})] + \zeta_0^{(i)}. \quad (27)$$

Continuing in the same logic as before, we let $\tilde{b}_n = \tilde{b}_\ell^{(i)}$, $\mathbf{g}_n = [\mathbf{A}]_i$ and $\mathbf{h}_n = [\mathbf{F}(t_\ell^{(i)})]$.

We then obtain

$$\tilde{b}_n = \text{vec}(\mathbf{g}_n \mathbf{h}_n^T)^T \text{vec}(\mathbf{C}(\mathbf{x})) + \zeta_0^{(i_n)}. \quad (28)$$

To keep things in matrix form, we first denote \mathbf{R} to be a matrix with rows $\mathbf{r}_n = \text{vec}(\mathbf{g}_n \mathbf{h}_n^T)^T$ and then denote $\tilde{\mathbf{R}}$ to be a matrix with row vectors $\tilde{\mathbf{r}}_n = [\mathbf{r}_n, e_{i_n}]$ where e_{i_n} is a length- I row vector with a 1 in the column corresponding to the machine that generated measurement n and zeros otherwise. We also denote the column vector $\tilde{\mathbf{C}}(\mathbf{x}) = [\text{vec}(\mathbf{C}(\mathbf{x}))^T, \zeta_0^{(1)}, \zeta_0^{(2)}, \dots, \zeta_0^{(I)}]^T$.

The measurement therefore satisfies

$$\tilde{b}_n = \tilde{\mathbf{r}}_n \tilde{\mathbf{C}}(\mathbf{x}). \quad (29)$$

For the system to be invertible we need $J \times (K + 1)$ of the vectors $\tilde{\mathbf{r}}_n$ to be linearly independent.

If we satisfy the condition set by Theorem 1 in (8), we also satisfy the condition set by Lemma 4 in (22). This means that there are JK rows \mathbf{r}_n of \mathbf{R} that are linearly independent. Now let us consider the extension $\tilde{\mathbf{R}}$, the corresponding JK rows from $\tilde{\mathbf{R}}$ will still be linearly independent (otherwise we reach a contradiction).

According to the assumptions of the corollary, $\tilde{\mathbf{R}}$ has, in addition to the JK rows already mentioned, one extra row $\tilde{\gamma}^{(i)}$ coming from each TEM⁽ⁱ⁾.

We would therefore like to check if there exist $JK + I$ coefficients $p_\ell^{(i)}$ such that

$$\sum_{i=1}^I \sum_{\ell=1}^{n_{\text{spikes}}^{(i)}} p_\ell^{(i)} \tilde{\gamma}_\ell^{(i)} = 0$$

$$\sum_{i=1}^I \sum_{\ell=1}^{n_{\text{spikes}}^{(i)}} p_{\ell}^{(i)} \begin{bmatrix} r_{\ell}^{(i)} \\ e_i \end{bmatrix} = 0 \quad (30)$$

Because the e_i 's are orthogonal, the above constraint can be translated to:

$$\begin{aligned} \sum_{\ell=1}^{n_{\text{spikes}}^{(i)}} p_{\ell}^{(i)} r_{\ell}^{(i)} &= 0, \quad \text{and} \\ \sum_{\ell=1}^{n_{\text{spikes}}^{(i)}} p_{\ell}^{(i)} &= 0, \quad \forall i = 1, \dots, I \end{aligned} \quad (31)$$

Because $n_{\text{spikes}}^{(i)} \leq K + 1$ for $\text{TEM}^{(i)}$, and the first $n_{\text{spikes}}^{(i)} - 1$ rows $\tilde{r}_k^{(i)}$'s from $\text{TEM}^{(i)}$ have full rank equal to $\min(\text{spikes}^{(i)} - 1, K)$ (according to Lemma 4), there is at most one solution for the $p_{\ell}^{(i)}$'s. This solution will depend on the $r_{\ell}^{(i)}$'s which follow a continuous probability distribution (given that the $t_{\ell}^{(i)}$'s also follow a continuous probability distribution). Therefore the $p_{\ell}^{(i)}$'s almost surely don't satisfy the second condition of (31), and the measurement vectors of $\tilde{\mathbf{R}}$ are almost surely linearly independent, making the system uniquely invertible with probability one. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable feedback which helped improve the quality of this paper.

REFERENCES

- [1] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 1964–1980, Jun. 2021.
- [2] T. Delbrück, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2010, pp. 2426–2429.
- [3] S.-C. Liu and T. Delbrück, "Neuromorphic sensory systems," *Curr. Opin. Neurobiol.*, vol. 20, no. 3, pp. 288–295, 2010.
- [4] S.-C. Liu, B. Rueckauer, E. Ceolini, A. Huber, and T. Delbrück, "Event-driven sensing for efficient perception: Vision and audition algorithms," *IEEE Signal Process. Mag.*, vol. 36, no. 6, pp. 29–37, Nov. 2019.
- [5] G. Gallego *et al.*, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.
- [6] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbrück, "Retinomorph event-based vision sensors: Bioinspired cameras with spiking output," *Proc. IEEE*, vol. 102, no. 10, pp. 1470–1484, Oct. 2014.
- [7] D. Gontier and M. Vetterli, "Sampling based on timing: Time encoding machines on shift-invariant subspaces," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 1, pp. 63–78, 2014.
- [8] H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, "Need for speed: A benchmark for higher frame rate object tracking," *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1125–1134.
- [9] N. T. Thao and D. Rzepka, "Time encoding of bandlimited signals: Reconstruction by pseudo-inversion and time-varying multiplierless FIR filtering," *IEEE Trans. Signal Process.*, vol. 69, pp. 341–356, Dec. 2020, doi: [10.1109/TSP.2020.3043809](https://doi.org/10.1109/TSP.2020.3043809).
- [10] R. Alexandru and P. L. Dragotti, "Reconstructing classes of non-bandlimited signals from time encoded information," *IEEE Trans. Signal Process.*, vol. 68, pp. 747–763, Dec. 2019, doi: [10.1109/TSP.2019.2961301](https://doi.org/10.1109/TSP.2019.2961301).
- [11] S. Rudresh, A. J. Kamath, and C. S. Seelamantula, "A time-based sampling framework for finite-rate-of-innovation signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 5585–5589.
- [12] A. A. Lazar and L. T. Tóth, "Perfect recovery and sensitivity analysis of time encoded bandlimited signals," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 10, pp. 2060–2073, Oct. 2004.
- [13] A. A. Lazar, "Time encoding with an integrate-and-fire neuron with a refractory period," *Neurocomputing*, vol. 58, pp. 53–58, 2004.
- [14] H. G. Feichtinger and K. Gröchenig, "Theory and practice of irregular sampling," *Wavelets: Math. Appl.*, vol. 1994, pp. 305–363, 1994.
- [15] A. A. Lazar, "Multichannel time encoding with integrate-and-fire neurons," *Neurocomputing*, vol. 65, pp. 401–407, 2005.
- [16] K. Adam, A. Scholefield, and M. Vetterli, "Multi-channel time encoding for improved reconstruction of bandlimited signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 7963–7967.
- [17] A. A. Lazar and L. T. Tóth, "Time encoding and perfect recovery of bandlimited signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 6, pp. VI-709.
- [18] A. N. Burkitt, "A review of the integrate-and-fire neuron model: I homogeneous synaptic input," *Biol. Cybern.*, vol. 95, no. 1, pp. 1–19, 2006.
- [19] A. Aldroubi and H. G. Feichtinger, "Non-uniform sampling: Exact reconstruction from non-uniformly distributed weighted-averages," in *Proc. Wavelet Analysis: Twenty Years' Developments*, 2002, pp. 1–8.
- [20] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, 1996.
- [21] D. Florescu and D. Coca, "A novel reconstruction framework for time-encoded signals with integrate-and-fire neurons," *Neural Comput.*, vol. 27, no. 9, pp. 1872–1898, 2015.
- [22] P. Martínez-Nuevo, H.-Y. Lai, and A. V. Oppenheim, "Amplitude sampling," in *Proc. IEEE 54th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2016, pp. 17–22.
- [23] H.-Y. Lai, P. Martínez-Nuevo, and A. V. Oppenheim, "An iterative reconstruction algorithm for amplitude sampling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 4576–4580.
- [24] H. G. Feichtinger, J. C. Príncipe, J. L. Romero, A. S. Alvarado, and G. A. Velasco, "Approximate reconstruction of bandlimited functions for the integrate and fire sampler," *Adv. Comput. Math.*, vol. 36, no. 1, pp. 67–78, 2012.
- [25] A. A. Lazar, E. K. Simonyi, and L. T. Tóth, "An overcomplete stitching algorithm for time decoding machines," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 9, pp. 2619–2630, Oct. 2008.
- [26] S. Saxena and M. Dableh, "Analyzing the effect of an integrate and fire encoder and decoder in feedback," in *Proc. IEEE 53rd Conf. Decis. Control*, 2014, pp. 3821–3828.
- [27] K. Adam, A. Scholefield, and M. Vetterli, "Sampling and reconstruction of bandlimited signals with multi-channel time encoding," *IEEE Trans. Signal Process.*, vol. 68, pp. 1105–1119, Jan. 2020, doi: [10.1109/TSP.2020.2967182](https://doi.org/10.1109/TSP.2020.2967182).
- [28] K. Adam, A. Scholefield, and M. Vetterli, "Encoding and decoding mixed bandlimited signals using spiking integrate-and-fire neurons," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 9264–9268.
- [29] A. A. Lazar and E. A. Pnevmatikakis, "Video time encoding machines," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 461–473, Mar. 2011.
- [30] M. Pacholska, K. Adam, A. Scholefield, and M. Vetterli, "Matrix recovery from bilinear and quadratic measurements," 2020, *arXiv:2001.04933*.
- [31] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 937–945.
- [32] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [33] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [34] K. Zhong, P. Jain, and I. S. Dhillon, "Efficient matrix sensing using rank-1 gaussian measurements," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2015, pp. 3–18.
- [35] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [36] T. Tao, *An Introduction to Measure Theory*, vol. 126. Providence, RI, USA: American Mathematical Society, 2011.