Imperial College London

3 May 2023

Timing is everything: model-based and learning-based reconstruction methods for event-driven cameras

Pier Luigi Dragotti

Imperial College London

Motivation

- Energy-efficient sensing inspired by nature (integrate and fire like neurons)
- The pixels are independent and asynchronous
- Pixel “fires” when measuring light intensity changes
- Information stored: location of pixel that fired and time of when it fired

Videos taken from Inivation.com

Motivation (cont'd)

- New sensing technologies also lead to new sampling challenges
  - How can we embed information related to complex signals into the timing information of spikes?
  - Besides its theoretical implications, addressing this question will lead to new neuromorphic sensing devices
  - Can new sampling results inspire new end-to-end neural networks?

Imperial College London

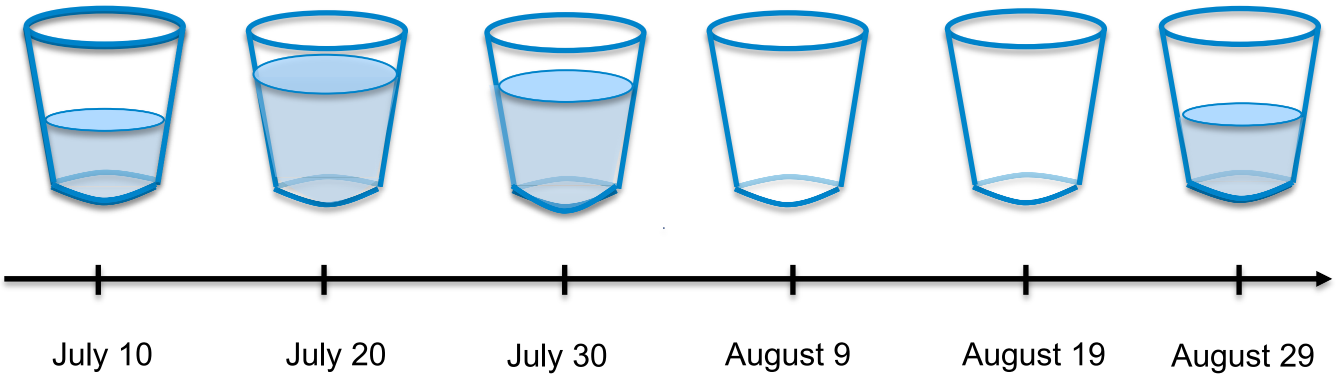
Outline

- Sampling based on timing
  - Integrate and fire systems
  - Time-based sampling of sparse signals (1D and 2D+t cases)
- Model-based deep learning for event cameras
  - End-to-end neural networks for event cameras
  - Deep unfolding approach for video reconstruction
- Conclusions and outlook

Imperial College London

Bio-Inspired Energy Efficient Sensing

- Current sensing methods are energy inefficient especially when low-latency is needed.
- Example: Rainfall estimation



Bio-Inspired Energy Efficient Sensing

Approach 2

- Only record the day when the bucket is full and then empty it

Imperial College London

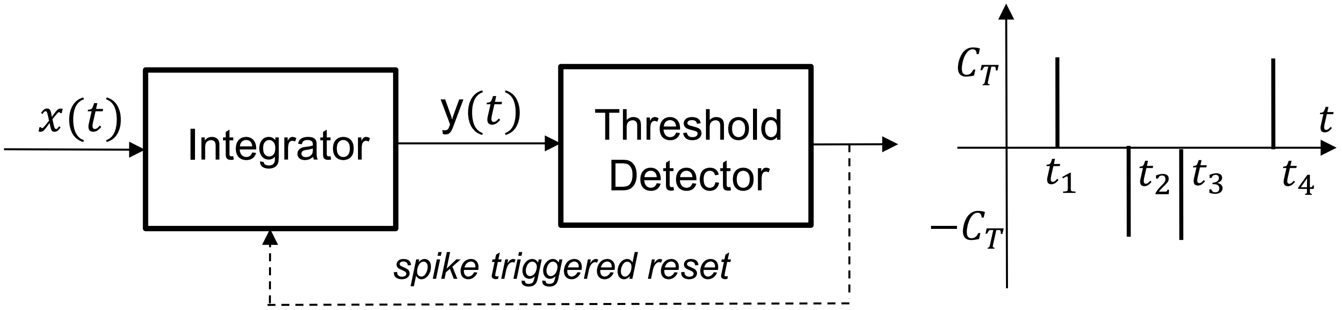
Bio-Inspired Energy Efficient Sensing

Approach 2 maps analogue information into a time sequence and is used by nature (e.g., integrate-and-fire neurons)

Time encoding appears in nature, as a mechanism used by neurons to represent sensory information as a sequence of action potentials, allowing them to process information **very efficiently**.

Time-Encoding Machines

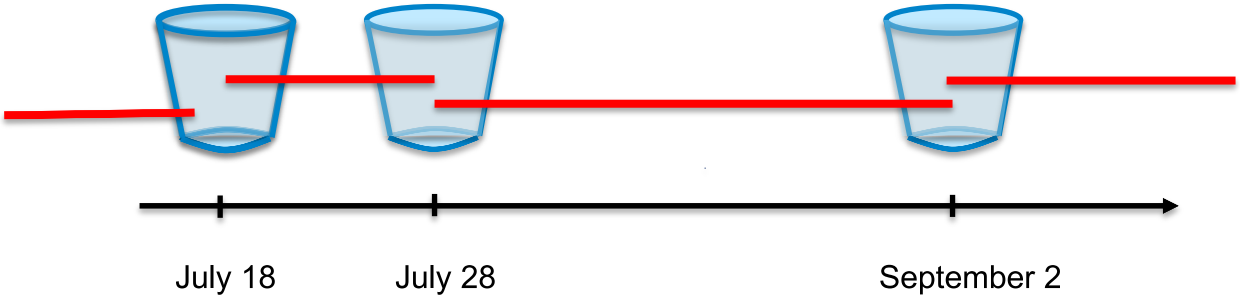
Integrate-and-fire System



Imperial College London

Reconstruction from time-encoded information

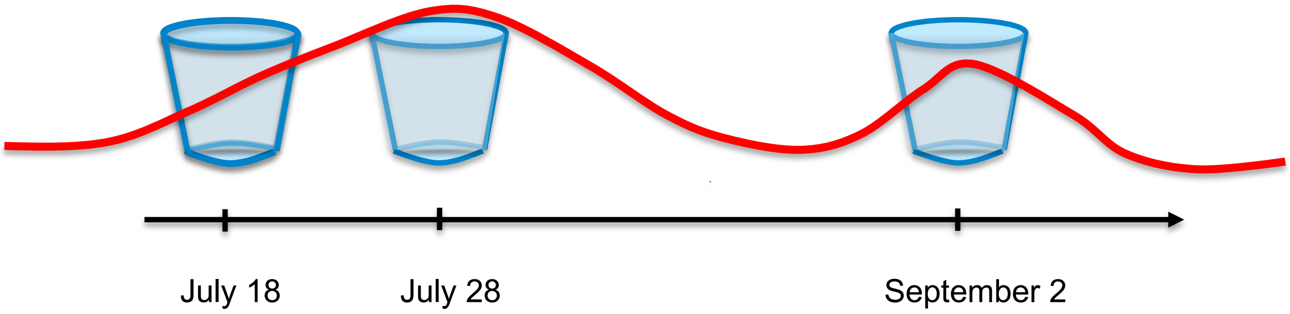
- Reconstruction achieved by imposing iteratively:
  - Consistency constraint
  - Signal prior (e.g., bandlimited function) constraint



Imperial College London

Reconstruction from time-encoded information

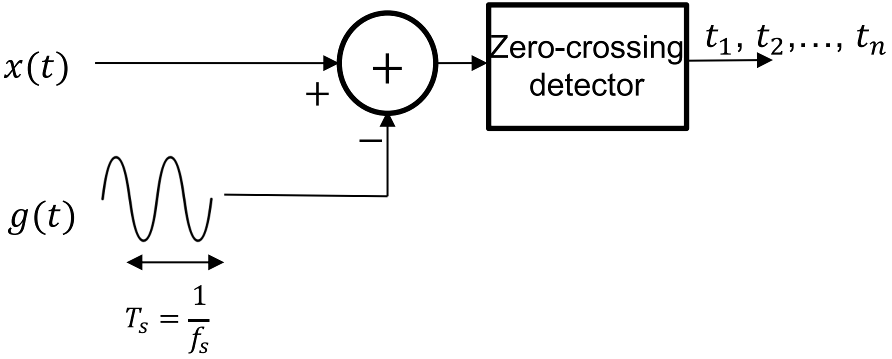
- Reconstruction achieved by imposing iteratively:
  - Consistency constraint
  - Signal prior (e.g., bandlimited function) constraint

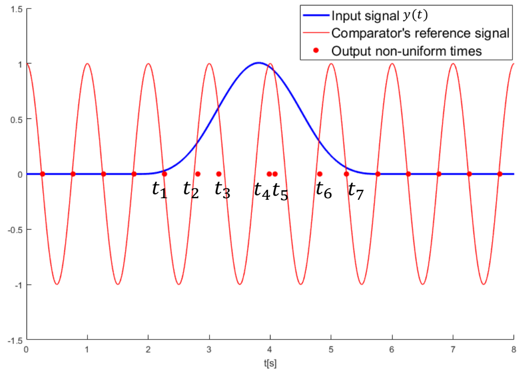


Imperial College London

Time-Encoding Machines

Comparator System



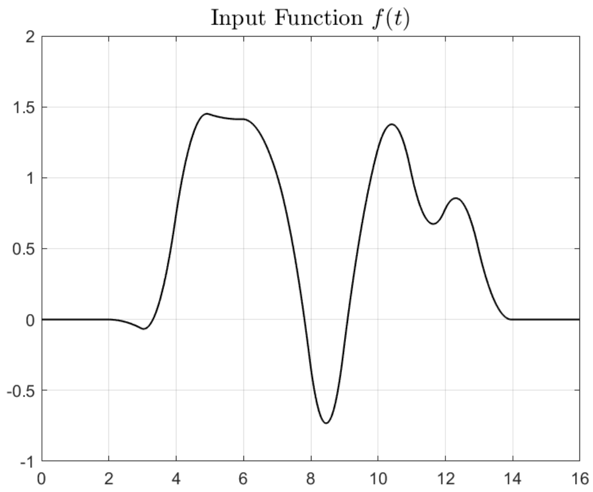


- At the crossing times, $x(t_n) - g(t_n) = 0$ hence $x(t_n) = g(t_n)$.

Imperial College London

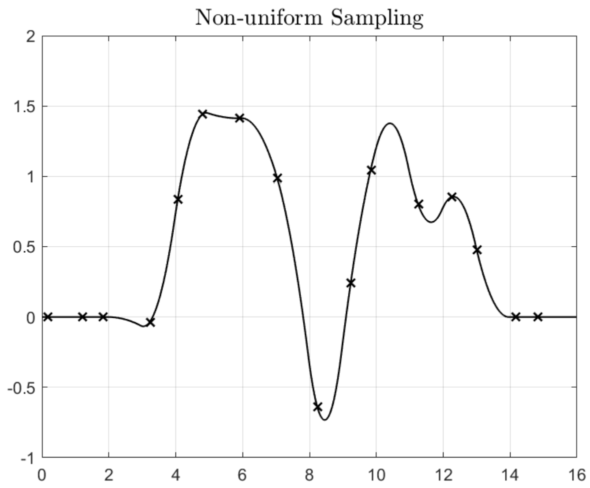
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



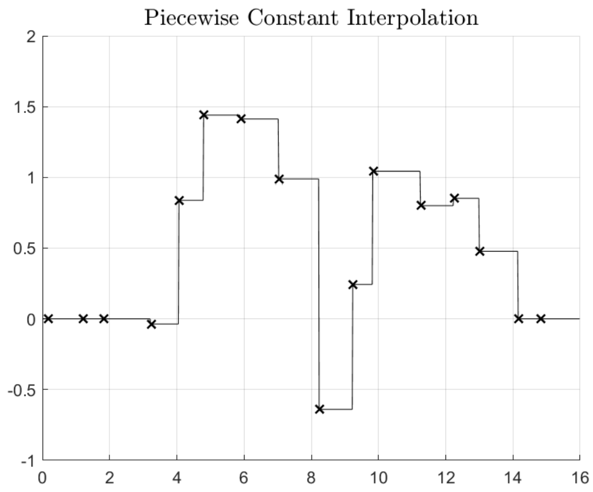
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



Reconstruction from time-encoded information

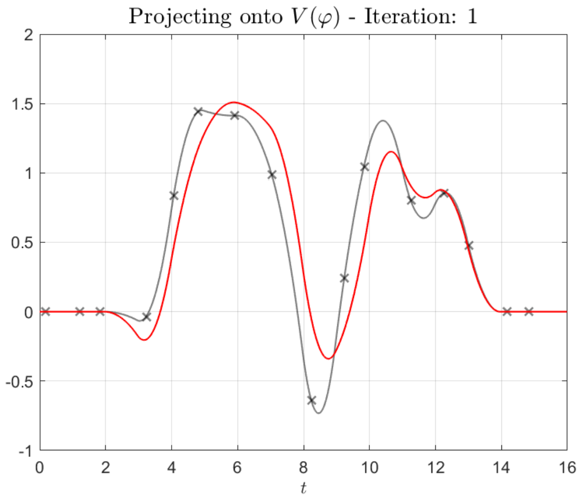
- The iterative approach proposed by Aldroubi and Grochenig



Imperial College London

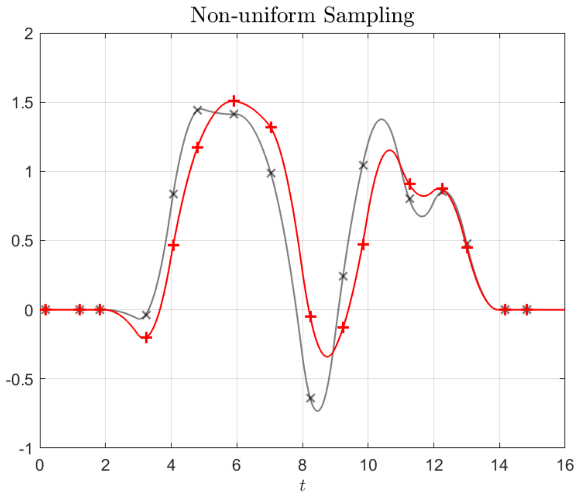
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



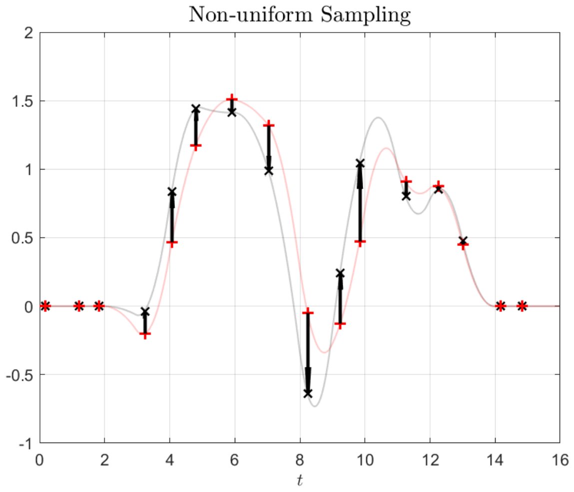
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



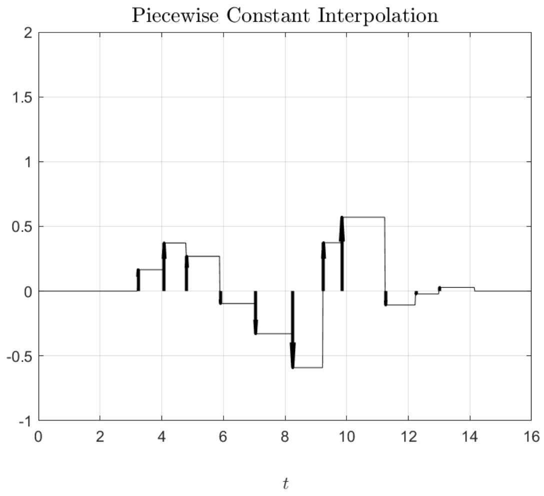
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



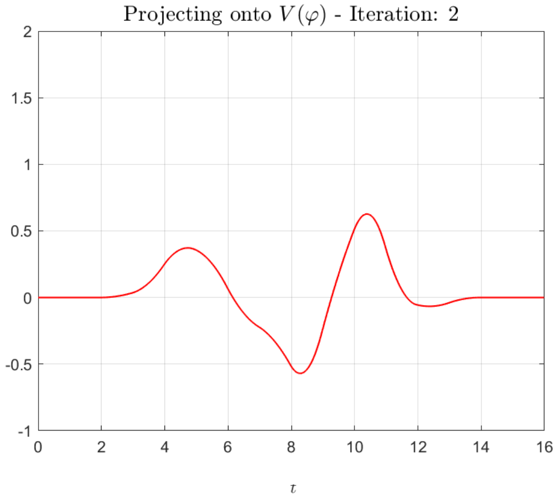
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



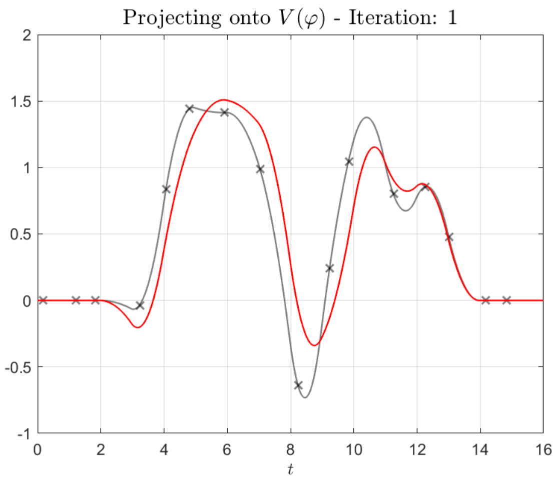
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig

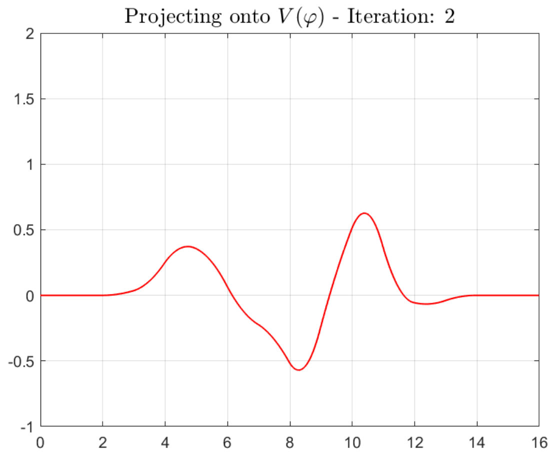


Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig

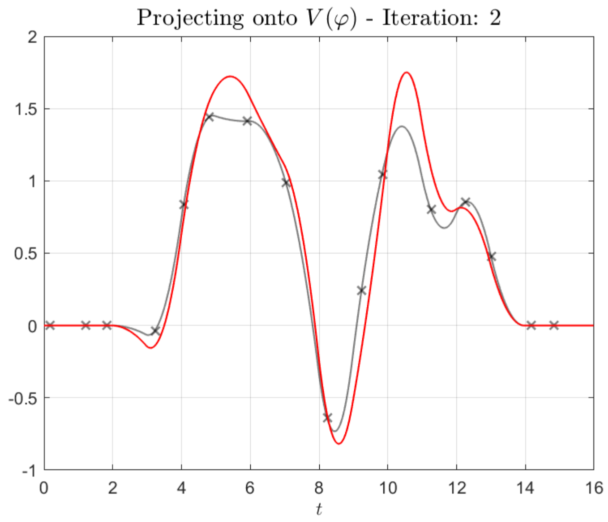


+



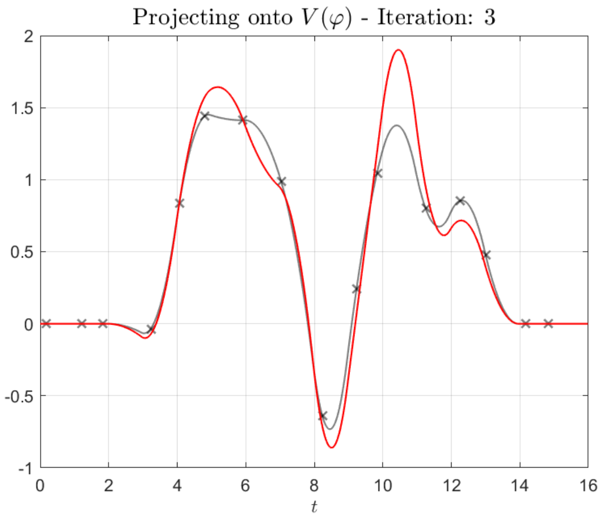
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



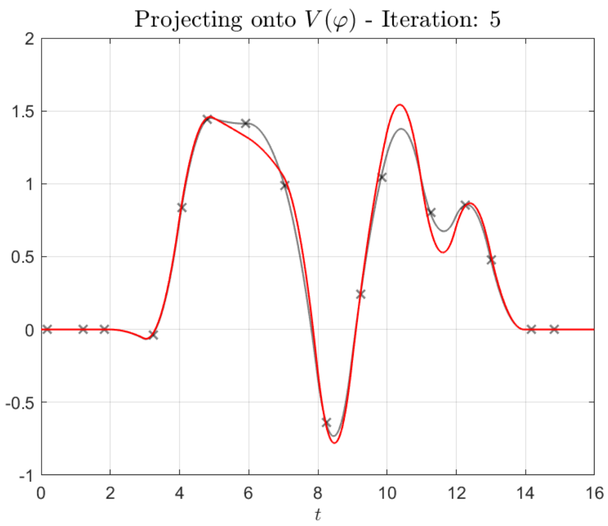
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



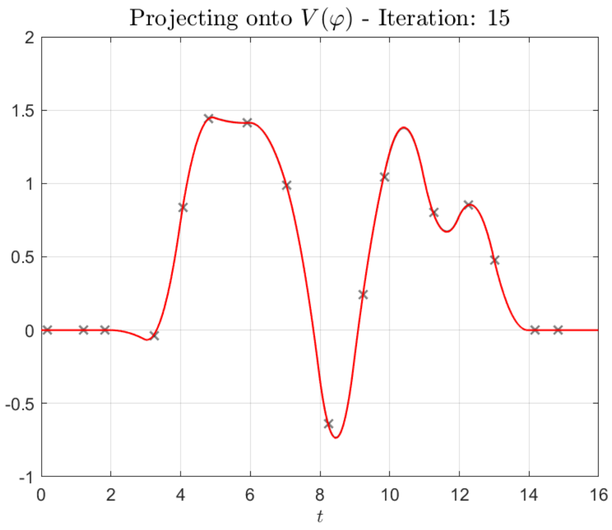
Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



Reconstruction from time-encoded information

- The iterative approach proposed by Aldroubi and Grochenig



Reconstruction from time-encoded information

- **Key result:** if the density of samples $D \geq 1$ then perfect reconstruction can be achieved (Aldroubi and Grochenig[2])

- **Key Issue 1**: In the case of uniform sampling the density is $D = 1$. This means that current TEMs are **less** energy efficient than uniform sampling!

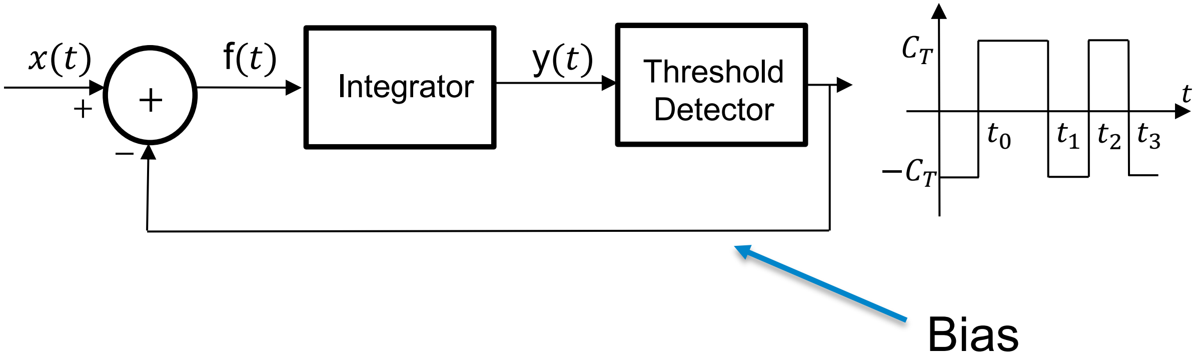
- **Key Issue 2:** Cannot sample sparse (non-bandlimited) signals with the current methods.

[2] A. Aldroubi and K, Grochenig, "Non-Uniform Sampling and Reconstruction in shift-invariant spaces" SIAM Review 2001

Imperial College London

Reconstruction from time-encoded information

- For integrate-and-fire machines exact reconstruction proved here: A. A. Lazar and L. T. Toth, “Time encoding and perfect recovery of bandlimited signals”, ICASSP 2003



See also: Gauntier-Vetterli-2014, Adam et al 2019,

Imperial College London

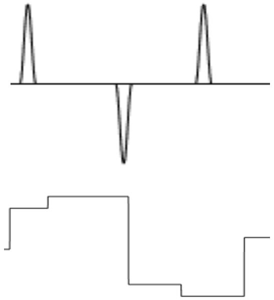
Time-based Sampling of Sparse Signals

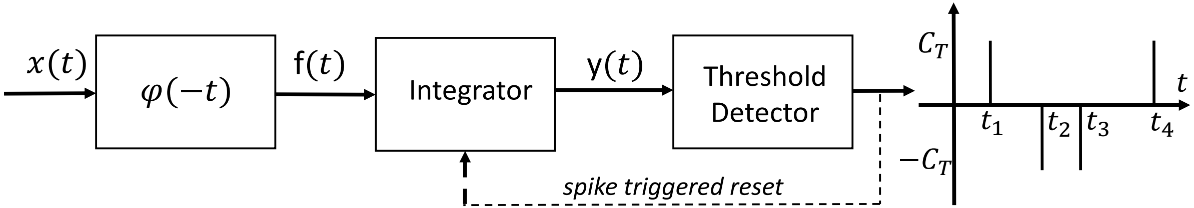
Signals:

- We consider sparse continuous-time signals like stream of pulses, piecewise constant or regular signals

Sensing Systems:

- We filter before using a TEM





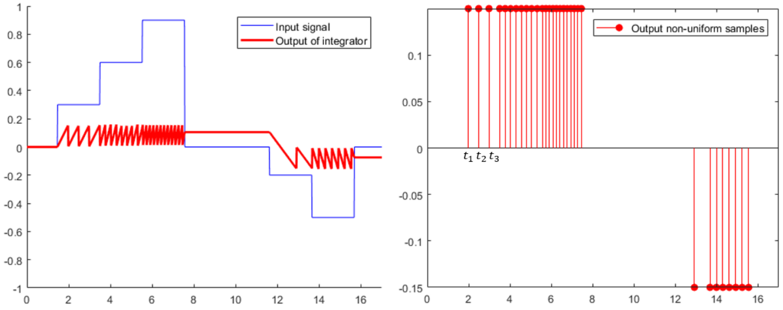
$$y(t_n) = \langle x(t), \varphi_n(t - t_n) \rangle$$

Imperial College London

Our approach for time decoding of signals

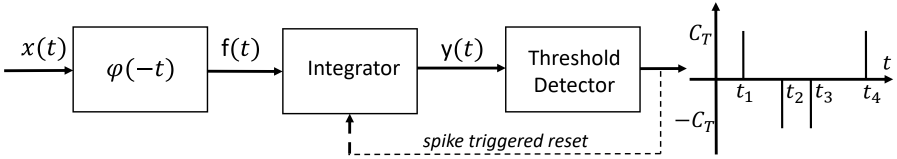
- Reconstruction of $x(t)$ depends on the
  - sampling kernel $\varphi(t)$
  - the density of time instants $\{t_n\}$

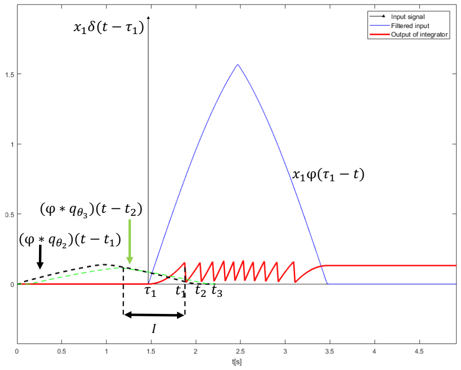
- We achieve a sufficient density of output samples by imposing conditions on:
  - The trigger mark of the integrator (**integrate-and-fire TEM**).



Imperial College London

Integrate and Fire TEM



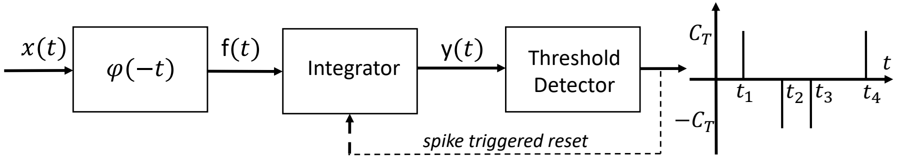


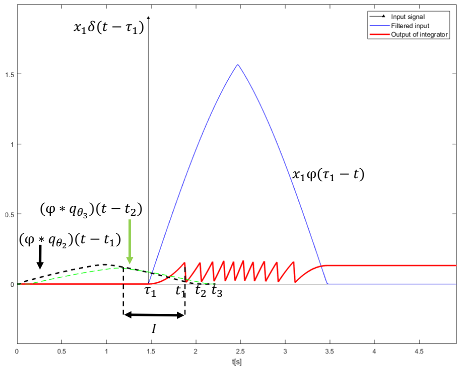
- Given the times $t_1, t_2, \ldots, t_n$, the amplitude values are

$$y_n = y(t_n) = \pm C_T = \int_{t_{n-1}}^{t_n} f(\tau)d\tau = \int_{t_{n-1}}^{t_n} \int x(\alpha)\varphi(\alpha - t)d\alpha d\tau.$$

Imperial College London

Integrate and Fire TEM





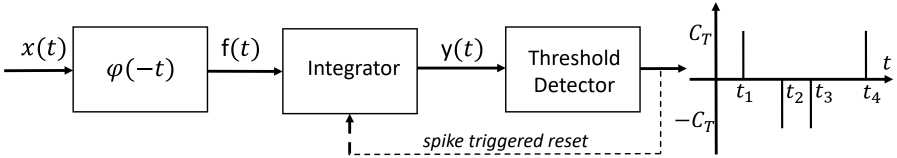
- Equivalently the output samples can be expressed as:

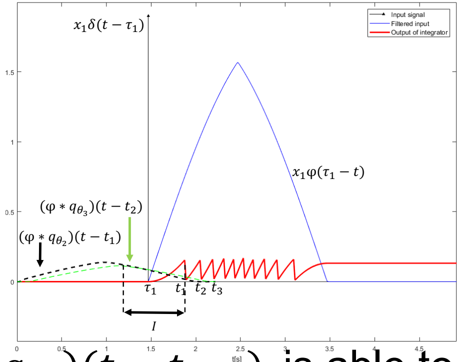
$$y(t_n) = \langle x(t), (\varphi * q_{\theta_n})(t - t_{n-1})\rangle,$$

where $\theta_n = t_n - t_{n-1}$ and $q_{\theta_n}(t)$ is defined as:

$$q_{\theta_n}(t) = \begin{cases} 1, & 0 \le t \le \theta_n, \\ 0, & otherwise. \end{cases}$$

Integrate and Fire TEM



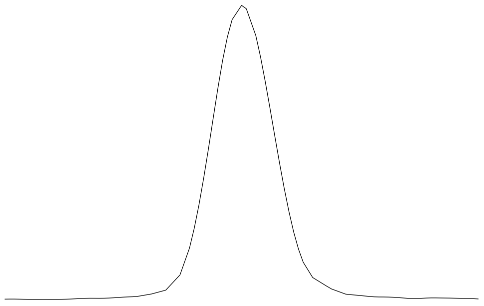


- When $\varphi(t)$ is e.g., an E-spline, the equivalent kernel $(\varphi * q_{\theta_n})(t - t_{n-1})$ is able to reproduce exponentials
- So trigger mark must guarantee enough samples in a short interval
- *Proposition:* when $C_T < \frac{A_{min}}{4\omega_0^2}\left(1 - \cos\left(\frac{\omega_0 L}{2}\right)\right)$ then $t_1, t_2, t_3 \in \left[\tau_1, \tau_1 + \frac{L}{2}\right]$ and perfect reconstruction is possible

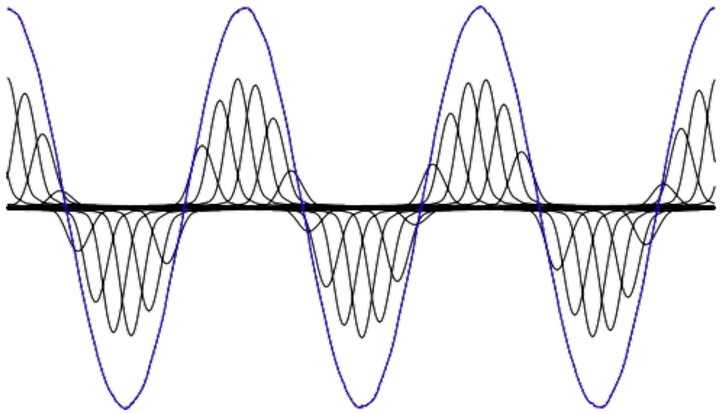
Imperial College London

Reproduction of Exponentials

$$\sum_{n} c_{m,n}\varphi(n-t) \approx e^{j\omega_m t}$$



Pulse shape



Reproduction of exponentials

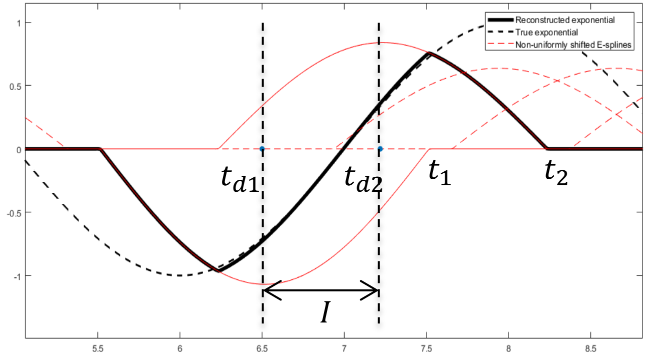
Imperial College London

Reproduction of Exponentials

- **Key Insight**: Reproduction of exponentials can be achieved locally in I, using at least two non-uniform shifts of the kernel:

$$\sum_{n=1}^{N} c_{m,n}\varphi_n(t-t_n) = e^{-\alpha_m t}, N \geq 2$$

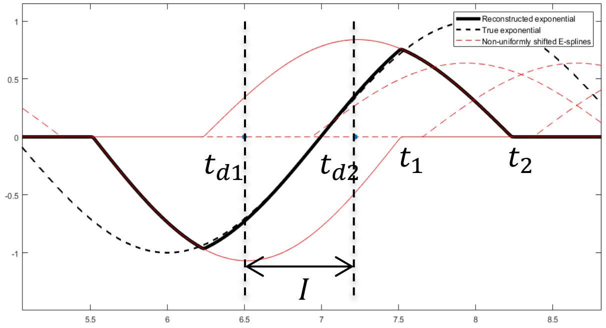
- The kernels should be continuous within that local interval I.



t_{d1} - discontinuity of $\varphi(t-t_1)$

t_{d2} - discontinuity of $\varphi(t-t_2)$

Reconstruction of an input Dirac from time-encoded information



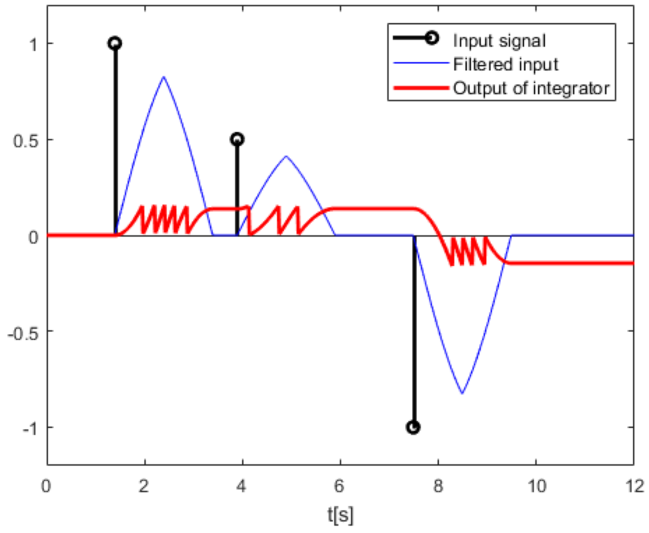
- The output samples are: $y(t_n) = \langle x(t), (\varphi * q_n)(t) \rangle = x_1 \varphi_n(\tau_1)$
- Since $\varphi_n(t) = a_{0,n} e^{\alpha_0 t} + a_{1,n} e^{\alpha_1 t}$, we find $\boldsymbol{c_1}$, $\boldsymbol{c_2}$, $\boldsymbol{d_1}$, $\boldsymbol{d_2}$ such that in $I_1 = [t_2 - T, t_1]$:

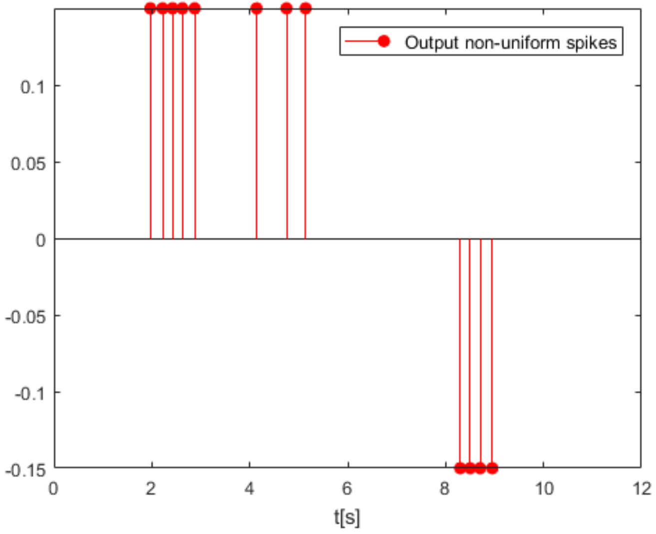
$$\boldsymbol{c_1}\ \varphi_1(t) + \boldsymbol{c_2}\ \varphi_2(t) = e^{\alpha_0 t}$$
$$\boldsymbol{d_1}\ \varphi_1(t) + \boldsymbol{d_2}\ \varphi_2(t) = e^{\alpha_1 t}$$

- We then use these coefficients to define the signal moments, in $I_1 = [t_2 - T, t_1]$:

$$s_0 = c_1 y(t_1) + c_2 y(t_2) = x_1 [c_1 \varphi_1(\tau_1) + c_2 \varphi_2(\tau_1)] = x_1 e^{\alpha_0 \tau_1}$$
$$s_1 = d_1 y(t_1) + d_2 y(t_2) = x_1 [d_1 \varphi_1(\tau_1) + d_2 \varphi_2(\tau_1)] = x_1 e^{\alpha_1 \tau_1}$$

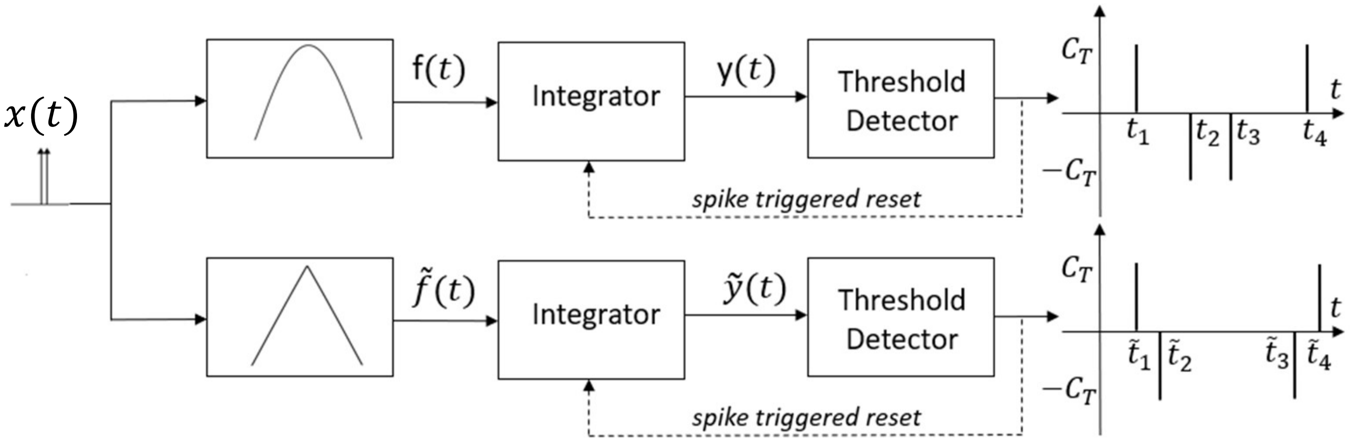
Integrate and Fire – Reconstruction of Pulses





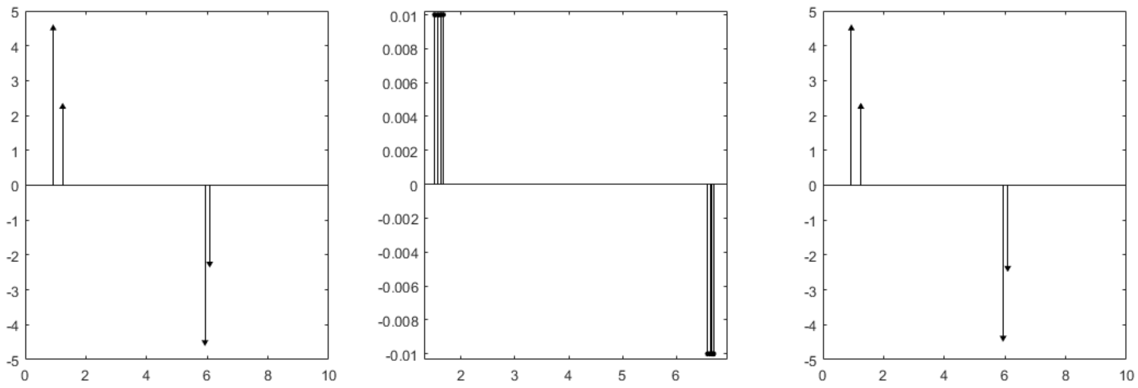
Imperial College London

Reconstruction of close pulses



Imperial College London

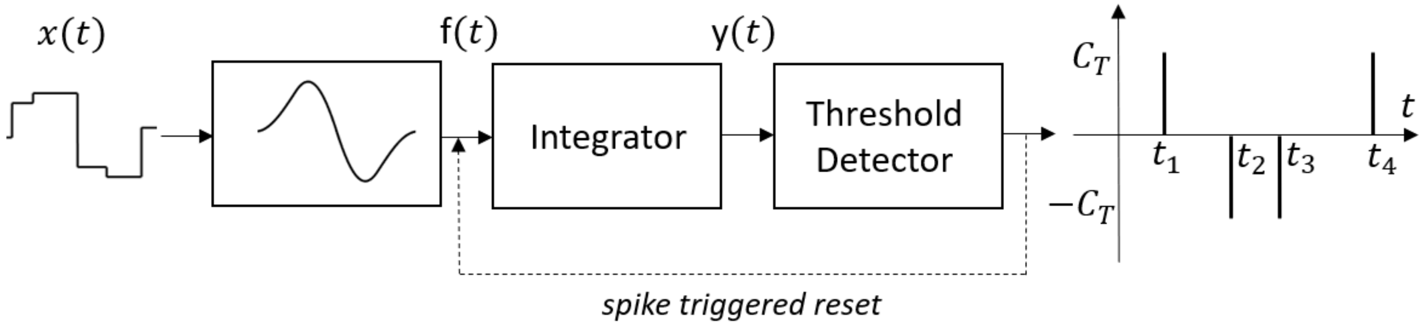
Integrate and Fire – Reconstruction of Pulses



(a) (b) (c)

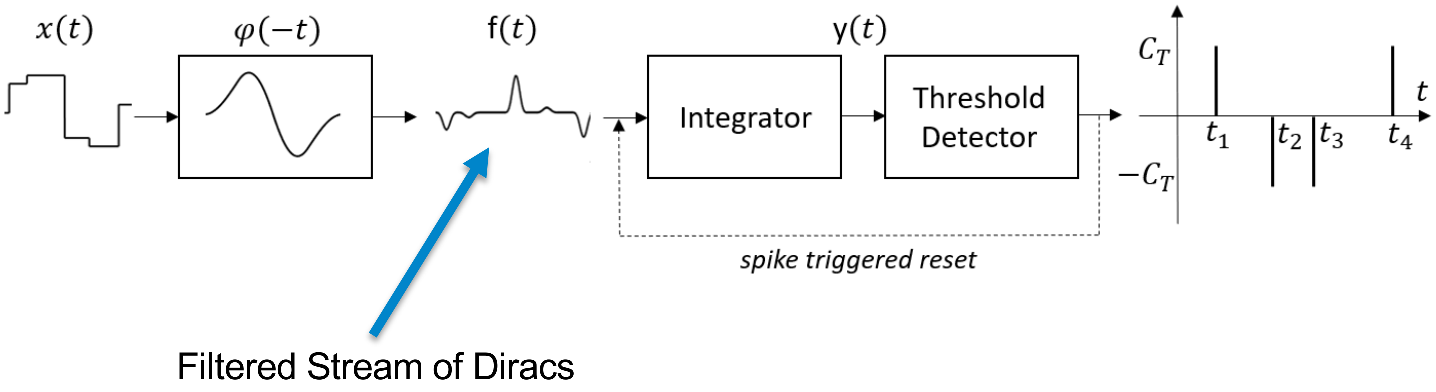
Imperial College London

Integrate and Fire – Piecewise Constant Signals



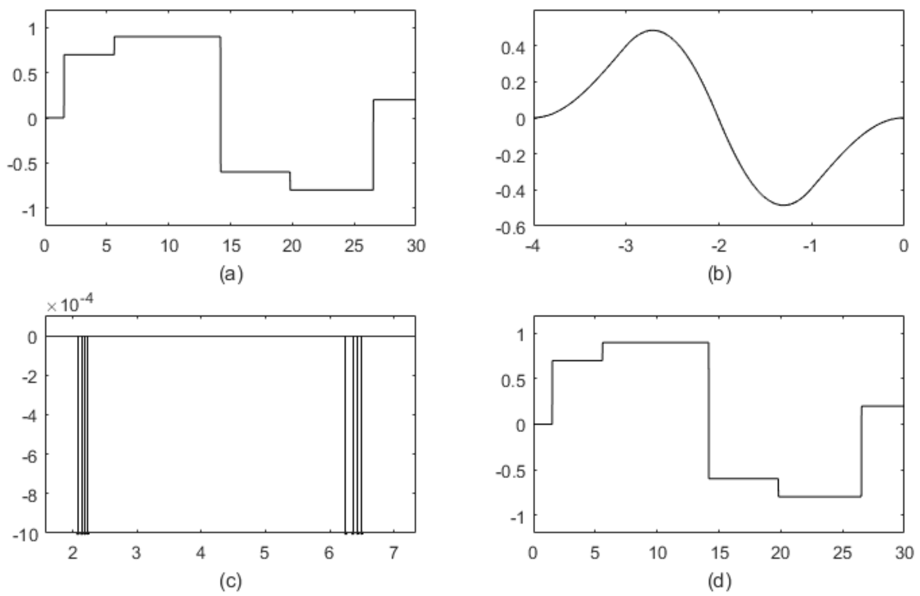
This is equivalent to the way a pixel operates in neuromorphic video cameras

Integrate and Fire – Piecewise Constant Signals



Imperial College London

Energy Efficient Sampling -Results

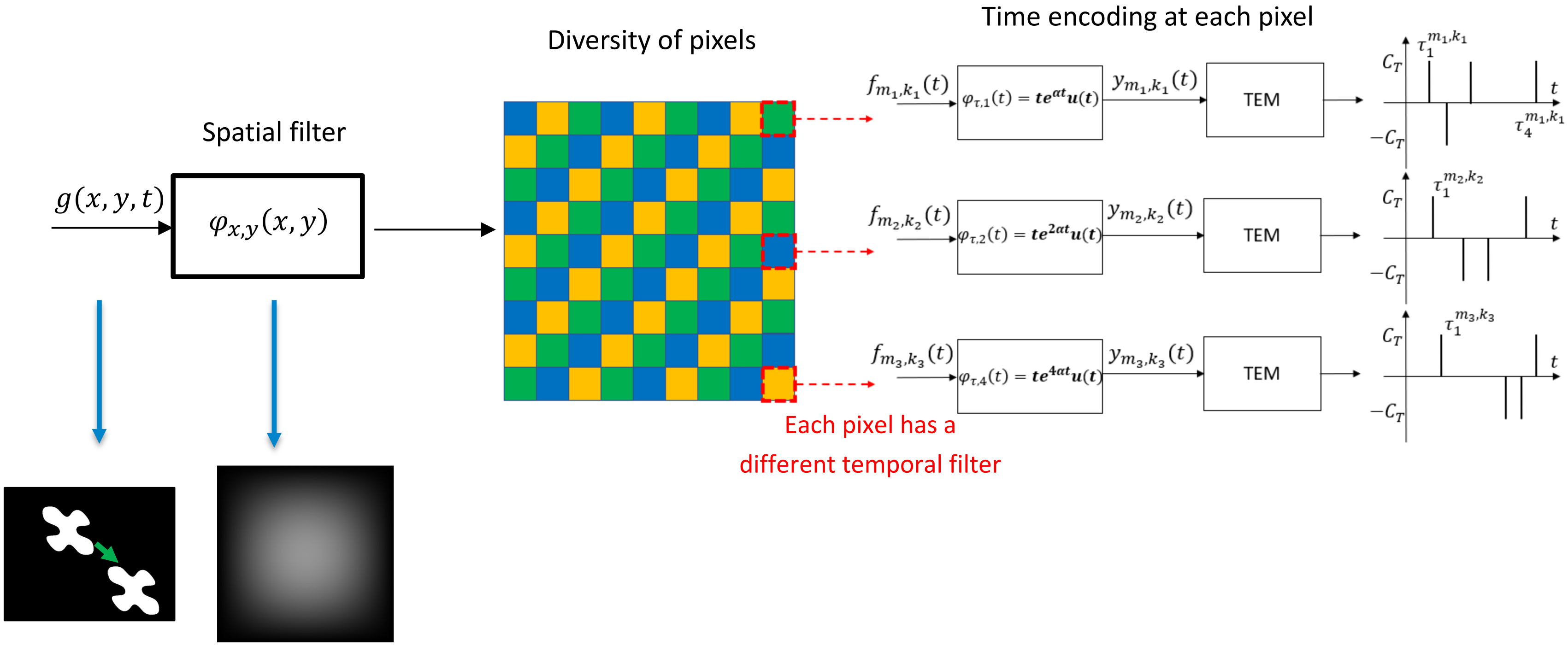


If the distance S between discontinuities is on average $S > (L-1)T$ with T being the sampling period in uniform sparse sampling then the new time encoding framework[3] is **more efficient** than sparse sampling (lower sampling density

[3]R. Alexandru and P.L. Dragotti, Reconstructing Classes of Non-bandlimited Signals from Time Encoded Information, IEEE Trans. on Signal Processing, vol.68, 2020.

Imperial College London

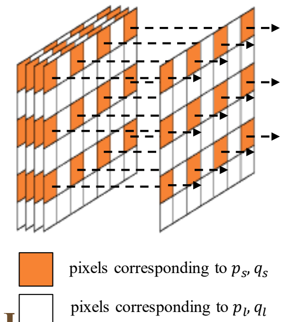
Integrate and Fire and Neuromorphic Cameras

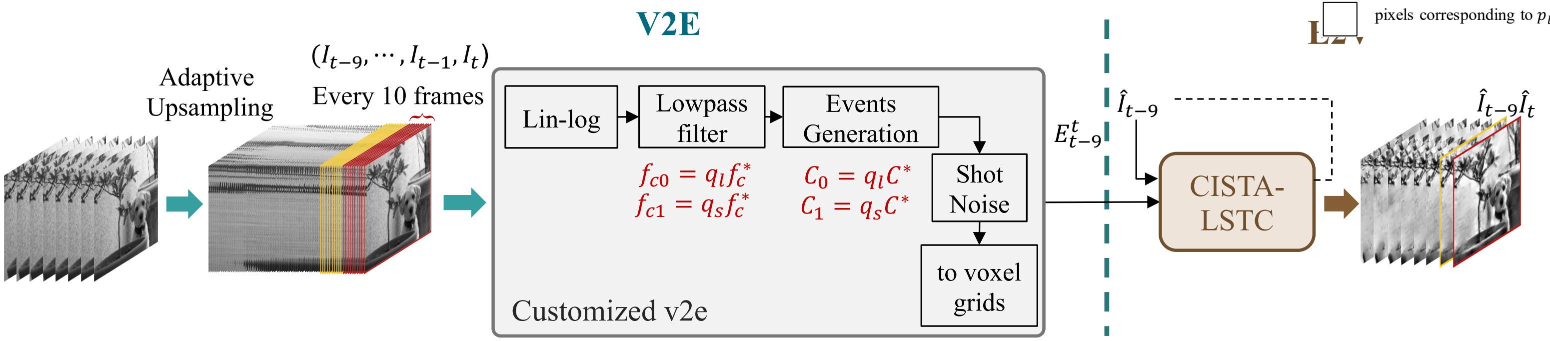


Imperial College London

Sensing Diversity for Neuromorphic Cameras

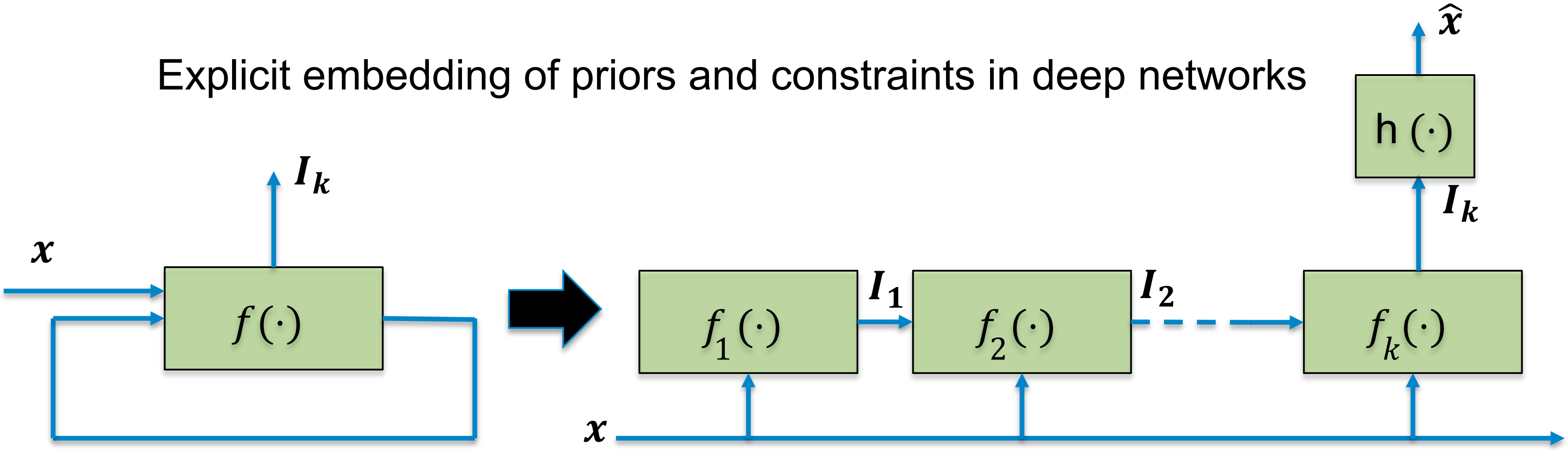
- Key insight: design an end-to-end neural network where the acquisition process is part of the learned architecture
- Key approach: each pixel behaves differently
- The network architecture for reconstruction is model-based





Deep Unfolding Strategy

Explicit embedding of priors and constraints in deep networks



Iterative algorithm with $\boldsymbol{x}$ as input and $\boldsymbol{I}$ as output

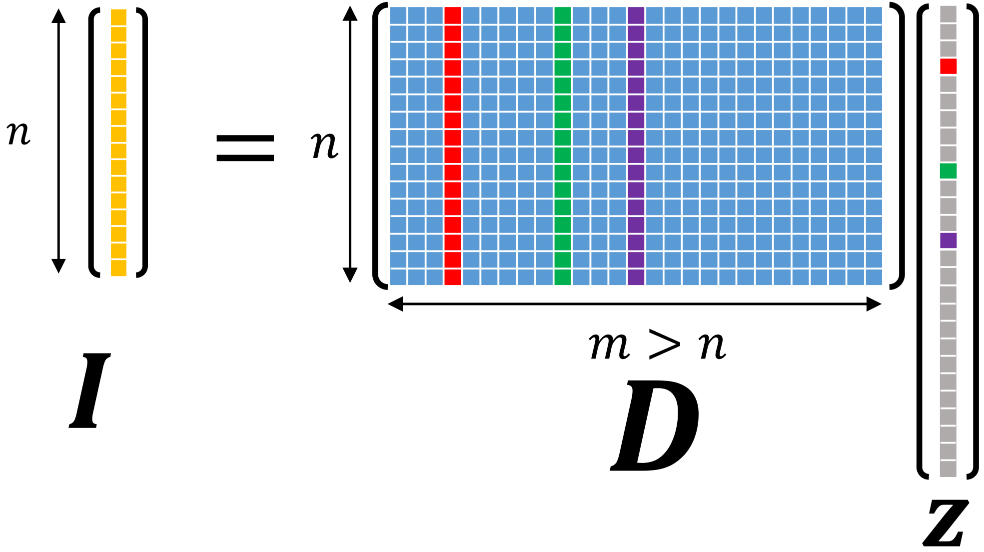
Unfolded version of the iterative algorithm with learnable parameters

Need to re-synthesize the input, if self-supervised

Imperial College London

Dictionary Learning

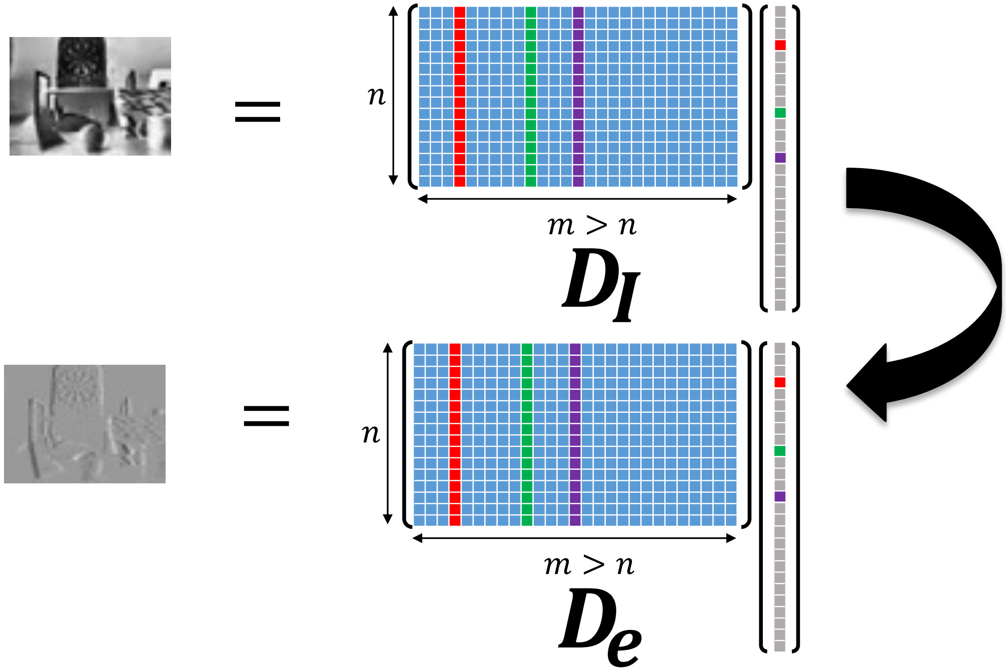
- The network architecture for reconstruction is model-based: intensity and event frames share the same sparse representation
- The dictionary is usually learned



Imperial College London

Model of dependency between intensity and events

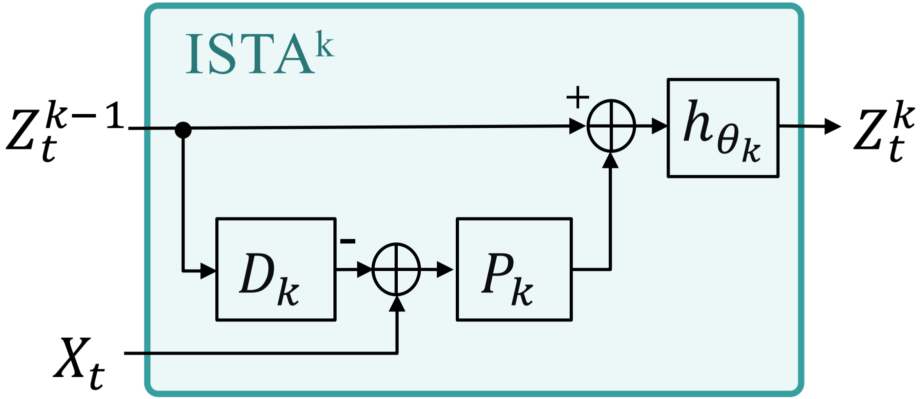
Assumption: intensity and event frames share the same sparse representation



Imperial College London

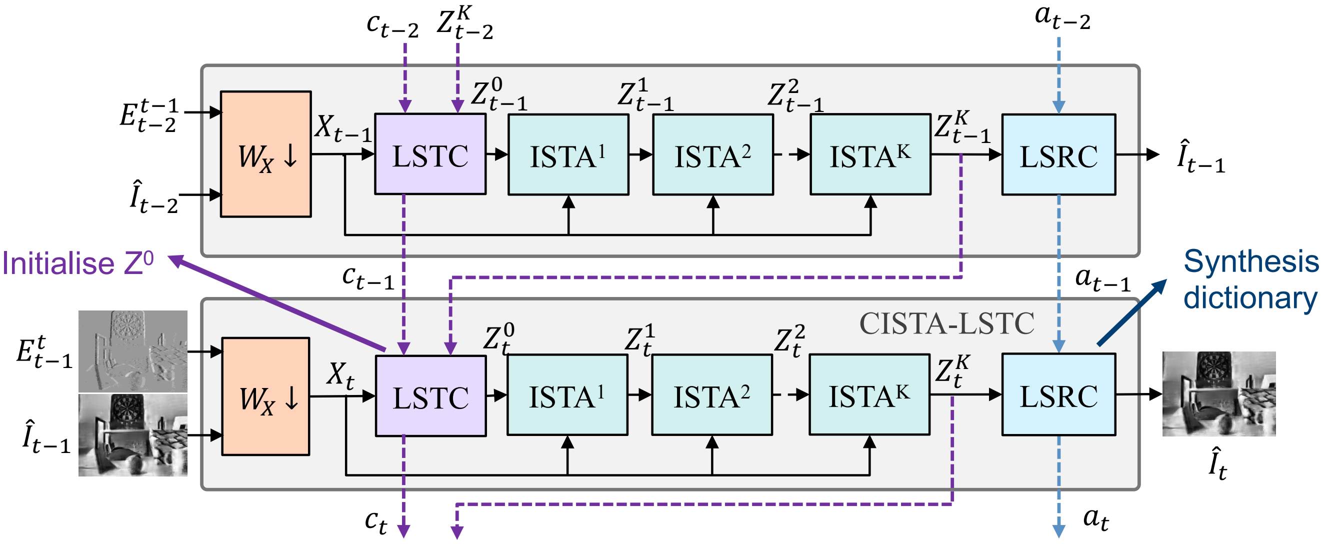
Deep Unfolding Strategy

- The network architecture for reconstruction is model-based: intensity and event frames share the same sparse representation Z_t
- The sparse vector can be found using ISTA: $Z_t^k = h_\theta(Z_t^{k-1} + D_k^T(X_t - D_k Z_t^{K-1})$



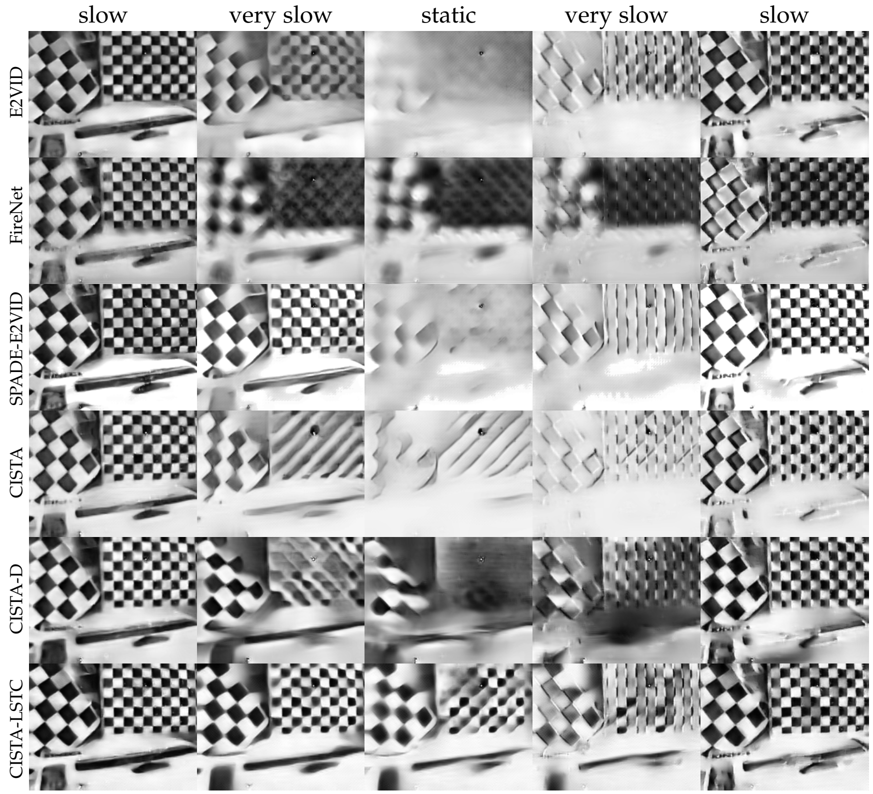
Imperial College London Model-based reconstruction from events

- The network architecture for reconstruction is model-based: intensity and event frames share the same sparse representation



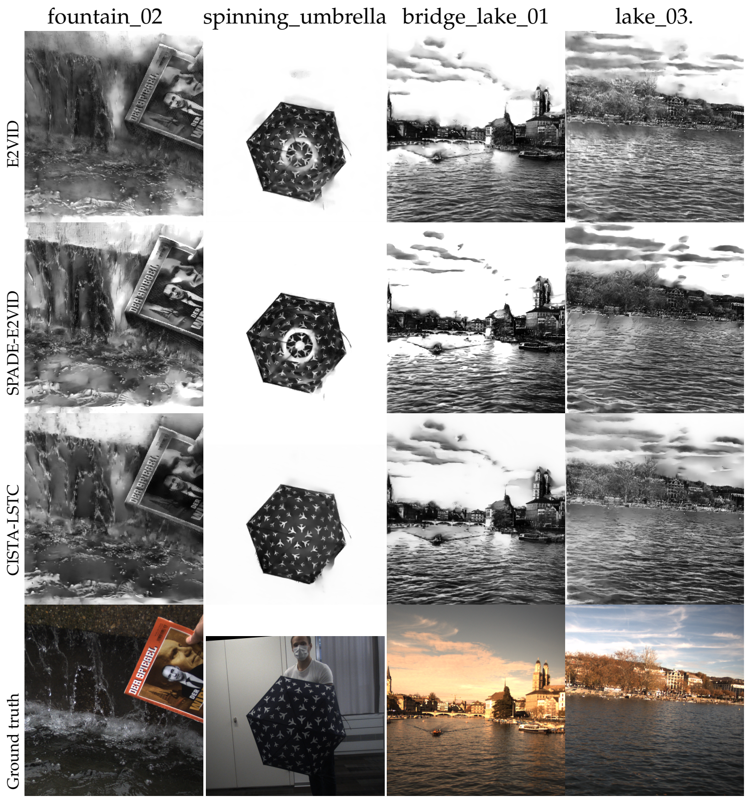
Imperial College London

Model-based reconstruction from events



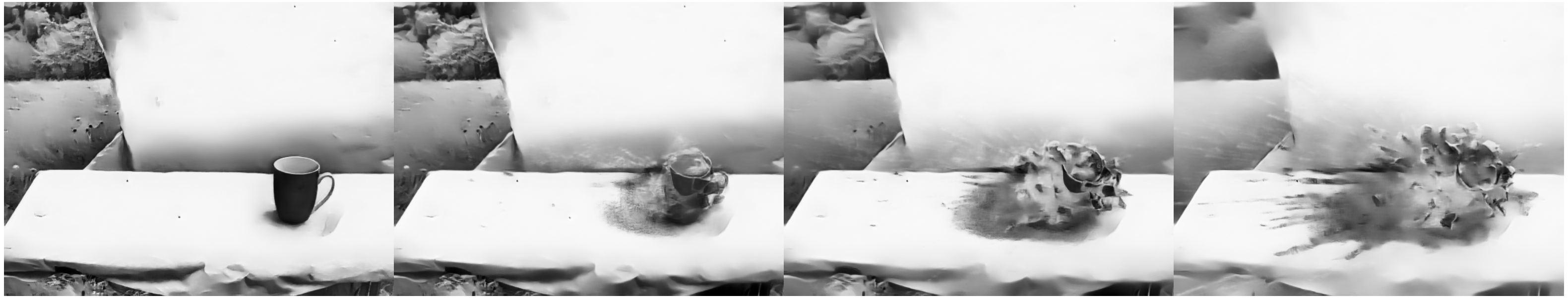
Imperial College London

Model-based reconstruction from events



Model-based reconstruction from events

gun_bullet_mug



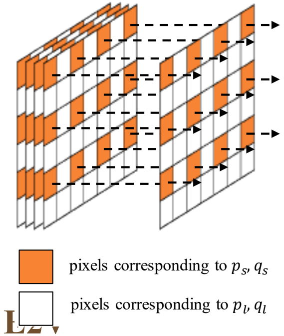
Imperial College London

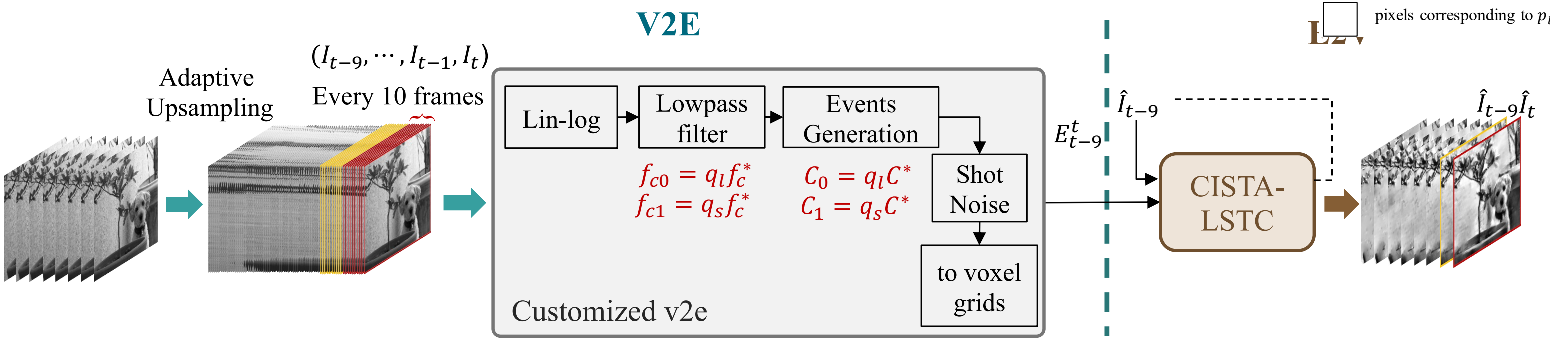
Model-based reconstruction from events

Imperial College London

Sensing Diversity for Neuromorphic Cameras

- Key insight: design an end-to-end neural network where the acquisition process is part of the learned architecture
- Key approach: each pixel behaves differently
- The network architecture for reconstruction is model-based





Imperial College London

Sensing Diversity for Neuromorphic Cameras


without sensing diversity


with sensing diversity

Imperial College London

Conclusions

- Neuromorphic sensing systems inspire a new paradigm for sampling
- Sampling provides insights into the design of event-driven systems (end-to-end learning)
- Model-based deep learning leads to lighter and more universal architectures

Imperial College London

Thank you!

Imperial College London

References

On time-based Sampling

- R. Alexandru and P. L. Dragotti, "Reconstructing classes of non-bandlimited signals from time encoded information", IEEE Transactions on Signal Processing, Vol.68, pp. 747-763, Year 2020

On Model-Based Deep Learning for intensity Video Reconstruction

- S. Liu, R.Alexandru and P.L. Dragotti, “Convolutional ISTA Network with Temporal Consistency Constraints for Video Reconstruction from Event Cameras”, IEEE ICASSP 2022
- S. Liu and P.L. Dragotti, Sensing Diversity and Sparsity Models for Event Generation and Video Reconstruction from Events, submitted to IEEE Trans. on Pattern Recognition and Machine Intelligence, 2022