Supplementary Material
Towards HDR and HFR Video from Rolling-Mixed-Bit Spikings

Yakun Chang[3,4#†] Yeliduosi Xiaokaiti[1,2#] Yujia Liu[1,2] Bin Fan[5]
Zhaojun Huang[1,2] Tiejun Huang[1,2] Boxin Shi[1,2*]

[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[3] Institute of Information Science, School of Computer Science, Beijing Jiaotong University
[4] Beijing Key Laboratory of Advanced Information Science and Network Technology
[5] National Key Lab of General AI, School of Intelligence Science and Technology, Peking University

ykchang@bjtu.edu.cn, {yongqiye, huangzhaojun}@stu.pku.edu.cn
{yujia_liu, binfan, tjhuang, shiboxin}@pku.edu.cn

In the supplementary material, we provide details of $Q_{\rm c}$ and $Q_{\rm th}$ in Eqns. (1) and (2) (Sec. 3.1), detailed derivation of Eqn. (6) (Sec. 3.2), details of the validation on the analysis using real-captured spikings (Sec. 3.2), additional implementation details (Sec. 4.3), and additional results on synthetic and real-synthetic data (Sec. 5). We further provide a supplementary video to show the HDR and HFR videos corresponding to Figs. 1, 8, 9, 10, 13, and 14.

7. Details of $Q_{\rm c}$ and $Q_{\rm th}$

As mentioned in Sec. 3.1, $Q_{\rm c}(t)$ consists of three components: the accumulated electrons $Q_{\rm a}$ in the previous interval, the photo-generated electrons $Q_{\rm p}$, and the dark electrons $Q_{\rm d}$ accumulated in the current interval. $Q_{\rm a}(t)$ is determined by the readout spiking of the previous interval, and it can be expressed as

$$Q_{\rm a}(t) = \begin{cases} Q_{\rm c}(t-1), & \text{if } S^L(t-1) = 0, \\ 0, & \text{if } S^L(t-1) \geq 1 \text{ or } t = 0. \end{cases} \quad (14)$$

The probability distribution of $Q_{\rm p}(t)$ has been given by Eqn. (8) in the main paper. $Q_{\rm d}(t)$ is obtained by integrating the dark current $I_{\rm d}$ over the interval τ. We model $I_{\rm d}$ as a spatially correlated Gaussian distribution:

$$I_{\rm d} \sim \mathcal{N}(\mu_{\rm d}, \sigma_{\rm d}^2). \quad (15)$$

Due to the limited precision of the readout circuit, the readout value of $Q_{\rm d}$ exhibits small deviations from the actual value. Here, we assume the readout value is obtained by rounding the actual value. For other parameters with similar situations, we adopt the same assumption as $Q_{\rm d}$. Then, for any $k \in \mathbb{Z}$, we have

$$\begin{aligned} \mathbb{P}(Q_{\rm d}(t) = k) &= \mathbb{P}(k - \frac{1}{2} \leq I_{\rm d}\tau < k + \frac{1}{2}), \\ &= \int_{(k-0.5)/\tau}^{(k+0.5)/\tau} \frac{1}{\sqrt{2\pi}\sigma_{\rm d}} \exp(-\frac{(x-\mu_{\rm d})^2}{2\sigma_{\rm d}^2}) {\rm d}x. \end{aligned} \quad (16)$$

Without considering noise, we have $Q_{\rm th} = C \cdot V$, where C is the capacitance of the capacitor and V is the voltage on the capacitor. Real-world spikings are affected by noise, and the noise is generated by deviation of the capacitor capacitance C^S, the voltage deviation V^S when resetting the voltage, and the voltage deviation $V^{T_0}(t)$ caused by temperature. The three types of noise are independent of each other, and they all follow Gaussian distributions:

$$\begin{aligned} C^S &\sim \mathcal{N}(0, \sigma_{\rm cs}^2), \\ V^S &\sim \mathcal{N}(0, \sigma_{\rm vs}^2), \\ V^{T_0}(t) &\sim \mathcal{N}(0, \sigma_{{\rm vt}_0}^2). \end{aligned} \quad (17)$$

Thus, the noise-affected $Q_{\rm th}$ is

$$Q_{\rm th} = (C + C^S)(V + V^S + V^{T_0}). \quad (18)$$

Let

$$Q_{\rm r} = C \cdot V, \quad (19)$$

and

$$Q_{\rm b} = (C + C^S)(V^S + V^{T_0}) + C^S V, \quad (20)$$

[#] Equal contribution. [*] Corresponding author.
[†] Majority of this work was done at Peking University.

then $Q_{\mathrm{th}} = Q_{\mathrm{r}} + Q_{\mathrm{b}}$. As Q_{r} is a constant, the distribution of Q_{th} is determined by the distribution of Q_{b}. For any $Q \in \mathbb{Z}$, we have

$$\mathbb{P}(Q_{\mathrm{b}} = Q) = \mathbb{P}(Q - 0.5 \leq Q_{\mathrm{b}} < Q + 0.5). \tag{21}$$

Let $V_1 = V^S$ and $V_2 = V^{T_0}$.We can establish a mapping from $(Q_{\mathrm{b}}, V_1, V_2)$ to (C^S, V^S, V^{T_0}), and represent the corresponding function relationships as $h_1(\cdot)$, $h_2(\cdot)$, and $h_3(\cdot)$ respectively

$$\begin{aligned} C^S &= h_1(Q_{\mathrm{b}}, V_1, V_2) = \frac{Q_{\mathrm{b}} - C(V_1 + V_2)}{V + V_1 + V_2}, \\ V^S &= h_2(Q_{\mathrm{b}}, V_1, V_2) = V_1, \\ V^{T_0} &= h_3(Q_{\mathrm{b}}, V_1, V_2) = V_2. \end{aligned} \tag{22}$$

Let $f(c_s, v_s, v_{t_0})$ denote the joint probability density of the random variables (C^S, V^S, V^{T_0}). Since C^S, V^S, and V^{T_0} are independent of each other, $f(c_s, v_s, v_{t_0})$ can be expressed as

$$f(c_s, v_s, v_{t_0}) = G(c_s, \sigma^2_{\mathrm{cs}}) G(v_s, \sigma^2_{\mathrm{vs}}) G(v_{t_0}, \sigma^2_{\mathrm{vt_0}}), \tag{23}$$

where $G(x, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp(-\frac{x^2}{2\sigma^2})$.

Denote $l(q_{\mathrm{b}}, v_1, v_2)$ as the joint probability density of the random variables $(Q_{\mathrm{b}}, V_1, V_2)$, it can be expressed as

$$\begin{aligned} l(q_{\mathrm{b}}, v_1, v_2) &= f(c_s, v_s, v_{t_0}) |J(q_{\mathrm{b}}, v_1, v_2)|, \\ &= f(\frac{q_{\mathrm{b}} - C(v_1 + v_2)}{V + v_1 + v_2}, v_1, v_2) |J(q_{\mathrm{b}}, v_1, v_2)|, \end{aligned} \tag{24}$$

where $J(q_{\mathrm{b}}, v_1, v_2)$ is the *Jacobian determinant*:

$$J(q_{\mathrm{b}}, v_1, v_2) = \begin{vmatrix} \frac{\partial h_1}{\partial q_b} & \frac{\partial h_1}{\partial v_1} & \frac{\partial h_1}{\partial v_2} \\ \frac{\partial h_2}{\partial q_b} & \frac{\partial h_2}{\partial v_1} & \frac{\partial h_2}{\partial v_2} \\ \frac{\partial h_3}{\partial q_b} & \frac{\partial h_3}{\partial v_1} & \frac{\partial h_3}{\partial v_2} \end{vmatrix} = \frac{1}{V + v_1 + v_2}. \tag{25}$$

The probability density function of Q_{b} is

$$f_{Q_{\mathrm{b}}}(q_{\mathrm{b}}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} l(q_{\mathrm{b}}, v_1, v_2) \mathrm{d}v_1 \mathrm{d}v_2. \tag{26}$$

Then for any $Q' \in \mathbb{Z}$, the probability of Q_{th} being Q' is

$$\begin{aligned} \mathbb{P}(Q_{\mathrm{th}} = Q') &= \mathbb{P}(Q_{\mathrm{b}} = Q' - Q_{\mathrm{r}}) \\ &= \int_{Q' - Q_{\mathrm{r}} - 0.5}^{Q' - Q_{\mathrm{r}} + 0.5} f_{Q_{\mathrm{b}}}(q_{\mathrm{b}}) \mathrm{d}q_{\mathrm{b}}. \end{aligned} \tag{27}$$

8. Explanations of information inheritance

The concept of information inheritance of a spiking camera comes from one of its characteristics. Taking the relationship between two adjacent intervals as an example. The accumulated electrons will reserve to the next interval if no spiking is emitted in the current interval, thus increasing the probability of emitting a spiking in the next interval. After a spiking is emitted in the current interval, electrons in the next interval start accumulating from zero, which results in a decreased probability of emitting spikings. This mechanism of information transmission imposes mutual constraints between consecutive intervals, leading to a relatively stable number of emitted spikings and consequently lower variance compared to when intervals are independent of each other. The mathematical proof is given below.

We start with the definition of the variance of S_N^L. We have

$$\begin{aligned} \mathrm{Var}(S_N^L) &= \mathrm{Var}(\sum_{i=1}^{N} S_i^L) \\ &= \sum_{i=1}^{N} \left[\mathrm{Var}(S_i^L) + 2 \sum_{j=i+1}^{N} \mathrm{Cov}(S_i^L, S_j^L) \right]. \end{aligned} \tag{28}$$

Here, S_i^L represents the i-th spiking in a spiking sequence. For any $i, j \in [1, N]$ and $i < j$, we have

$$\mathrm{Cov}(S_i^L, S_j^L) = \mathbb{E}(S_i^L S_j^L) - \mathbb{E}(S_i^L)\mathbb{E}(S_j^L). \tag{29}$$

$\mathbb{E}(S_i^L S_j^L)$ can be expressed as

$$\begin{aligned} \mathbb{E}(S_i^L S_j^L) &= \sum_{a=0}^{L} \sum_{b=0}^{L} ab \mathbb{P}(S_i^L = a, S_j^L = b) \\ &= \sum_{a=1}^{L} \sum_{b=1}^{L} ab \mathbb{P}(S_i^L = a, S_j^L = b) \\ &= \sum_{a=1}^{L} \sum_{b=1}^{L} ab \mathbb{P}(S_j^L = b | S_i^L = a) \mathbb{P}(S_i^L = a). \end{aligned} \tag{30}$$

For $\mathbb{P}(S_j^L = b | S_i^L = a)$, the condition is such that, when the i-th interval emits a spiking, the accumulated electrons will reset to zero at the end of the i-th accumulation interval for any $a \geq 1$. This means that the result of this expression is independent of the specific value of a, and is equivalent to $P(S_j^L = b | S_i^L \neq 0)$. Then we have

$$\begin{aligned} \mathbb{E}(S_i^L S_j^L) &= \sum_{a=1}^{L} \sum_{b=1}^{L} ab \mathbb{P}(S_j^L = b | S_i^L \neq 0) \mathbb{P}(S_i^L = a) \\ &= \sum_{a=1}^{L} a \mathbb{P}(S_i^L = a) \cdot \sum_{b=1}^{L} b \mathbb{P}(S_j^L = b | S_i^L \neq 0) \\ &= \mathbb{E}(S_i^L) \mathbb{E}(S_j^L | S_i^L \neq 0). \end{aligned} \tag{31}$$

Then $\mathrm{Cov}(S_i^L S_j^L)$ can be expressed as

$$\begin{aligned} \mathrm{Cov}(S_i^L, S_j^L) &= \mathbb{E}(S_i^L)\mathbb{E}(S_j^L | S_i^L \neq 0) - \mathbb{E}(S_i^L)\mathbb{E}(S_j^L) \\ &= \mathbb{E}(S_i^L) \left[\mathbb{E}(S_j^L | S_i^L \neq 0) - \mathbb{E}(S_j^L) \right]. \end{aligned} \tag{32}$$

Now let's add up covariance terms:

$$\begin{aligned}&\sum_{j=i+1}^{N}\text{Cov}(S_i^L,S_j^L)\\ =&\mathbb{E}(S_i^L)\left[\mathbb{E}(\sum_{j=i+1}^{N}S_j^L|S_i^L\neq 0)-\mathbb{E}(\sum_{j=i+1}^{N}S_j^L)\right].\end{aligned} \tag{33}$$

In Eqn. (33), the $\mathbb{E}(\sum_{j=i+1}^{N}S_j^L)$ expression calculates the expected number of spikings emitted from the $(i+1)$-th to the Nth interval, where the initial electron count in the $(i+1)$-th interval is one of the values in $[0, Q/L)$ following a certain distribution. The condition $(S_i\neq 0)$ ensures that the initial electron count in the $(i+1)$-th interval is always zero. The $\mathbb{E}(\sum_{j=i+1}^{N}S_j^L|S_i^L\neq 0)$ expression is at a disadvantage in terms of initial electron count compared to $\mathbb{E}(\sum_{j=i+1}^{N}S_j^L)$, resulting in a smaller expected number of spikings emitted. Combining $\mathbb{E}(S_i^L)\geq 0$, we obtain $\sum_{j=i+1}^{N}\text{Cov}(S_i^L,S_j^L)\leq 0$. Substituting this inequality into Eqn. (28), we obtain

$$\text{Var}(S_N^L)\leq\sum_{i=1}^{N}\text{Var}(S_i^L)=N\cdot\text{Var}(S^L). \tag{34}$$

9. Detailed derivation of Eqn. (6)

The presence of noise has an influence on both Q_{th} and Q_{d}, subsequently affecting S^L. The probability distribution of S^L varies with different levels of noise. Here, the utilization of conditional probability is capable of solving the calculation of $P(S^L=H)$. For $\mathbb{P}(S^L=H|Q_{\text{th}}=Q',Q_{\text{d}}=k)$ and $\mathbb{P}(Q_{\text{th}}=Q',Q_{\text{d}}=k)$, we simply denote them as $\mathbb{P}_{S^L}(H|Q',k)$ and $\mathbb{P}(Q',k)$. According to *the law of total probability*, we have

$$\mathbb{P}(S^L=H)=\sum_{Q'=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\mathbb{P}_{S^L}(H|Q',k)\mathbb{P}(Q',k). \tag{35}$$

Since Q_{th} and Q_{d} are independent of each other, we have

$$\mathbb{P}(Q',k)=P_{Q'}P_k. \tag{36}$$

Now we derive the expression for $\mathbb{P}_{S^L}(H|Q',k)$. Define $\{S^L\}$ as a sequence of spikings readout from N intervals, and L is the quantization level of the spiking signal. We assume the number that occurs the spiking signal H is N_H, then

$$\mathbb{P}_{S^L}(H|Q',k)=\lim_{N\to\infty}\frac{N_H}{N}. \tag{37}$$

$\{S^L\}$ is composed of a series of accumulation sequences. The accumulation sequence is defined by the condition when a pixel begins to accumulate electrons with an initial state of zero accumulated electrons, after accumulating $n-1$ intervals, it finally emits h spikings. We then collect these signals to obtain an accumulation sequence of length n:

$$\{A_n^h\}=\underbrace{0,0,\cdots,0}_{n-1},h. \tag{38}$$

In the sequence $\{S^L\}$, the number of occurrences of $\{A_n^h\}$ is denoted as $R(\{A_n^h\})$. Then, the total length of the spiking sequence S^L is the sum of the product of $R(\{A_n^h\})$ and the length of all possible accumulation sequences:

$$N=\sum_{h=1}^{L}\sum_{n=1}^{\infty}n\,R(\{A_n^h\}). \tag{39}$$

The total number of accumulation sequences that emit H spikings represents the number of times H spikings are emitted in the S_n^L sequence, that is

$$N_H=\sum_{n=1}^{\infty}R(\{A_n^H\}). \tag{40}$$

Substituting the above two equations into Eqn. (37), we have

$$\mathbb{P}_{S^L}(H|Q',k)=\frac{\sum_{n=1}^{\infty}R(\{A_n^H\})}{\sum_{h=1}^{L}\sum_{n=1}^{\infty}n\,R(\{A_n^h\})}. \tag{41}$$

The proportion of $R(\{A_n^h\})$ in all accumulation sequences is the probability of its occurrence

$$\mathbb{P}(\{A_n^H\})=\frac{R(\{A_n^H\})}{\sum_{h=1}^{L}\sum_{n=1}^{\infty}R(\{A_n^h\})}, \tag{42}$$

which is equivalent to P_n^H in Eqn. (7).

Divide the numerator and denominator of Eqn. (41) by $\sum_{h=1}^{L}\sum_{n=1}^{\infty}R(\{A_n^h\})$, we get

$$\begin{aligned}\mathbb{P}_{S^L}(H|Q',k)=&\frac{\sum_{n=1}^{\infty}\mathbb{P}(\{A_n^H\})}{\sum_{h=1}^{L}\sum_{n=1}^{\infty}n\,\mathbb{P}(\{A_n^h\})}\\ =&\frac{\sum_{n=1}^{\infty}P_n^H}{\sum_{h=1}^{L}\sum_{n=1}^{\infty}nP_n^h}.\end{aligned} \tag{43}$$

By substituting equation Eqns. (36) and (43) into Eqn. (35), we get Eqn. (6).

10. Validation on real-captured spikings

To validate the correctness of Eqn. (4) and Eqn. (5), we calculate the expectation and variance of real-captured spikings. For a spiking sequence $\{S^L\}$ of length N, we normalize the spiking readout values to $[0,1]$, and then the expectation of the normalized spiking sequence is driven by

$$\mu=\frac{1}{N}\sum_{i=1}^{N}\frac{S_i^L}{L}, \tag{44}$$

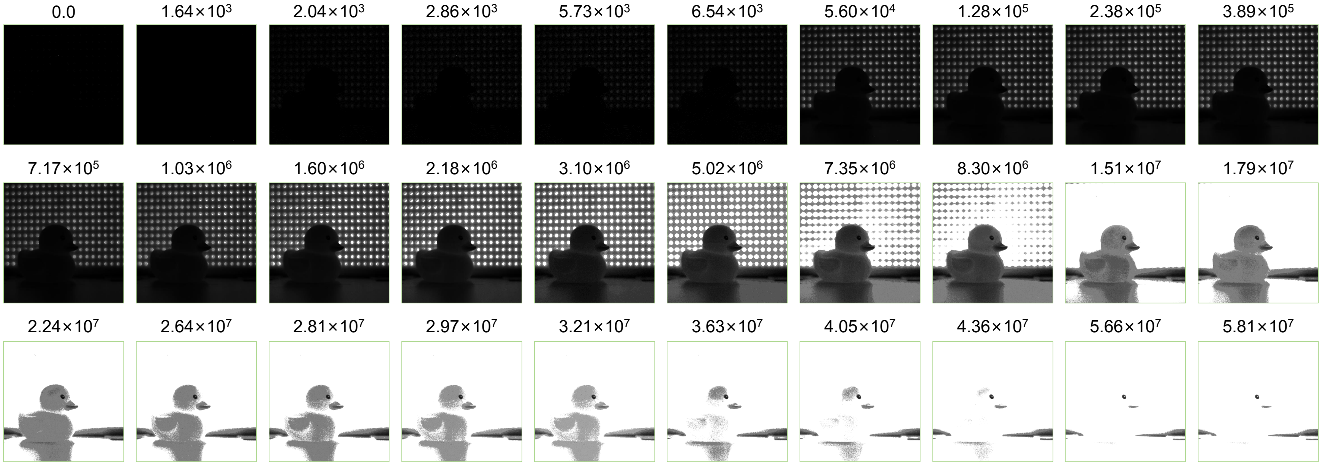

Figure 11. 30 images taken by a spiking camera under varying exposure conditions of the same scene. We use Eqn. (44) to obtain the average value of each pixel to reconstruct the image. The number above each image is the average exposure for all pixels. The unit of exposures is photon$/(\mu m^2 \cdot s)$.

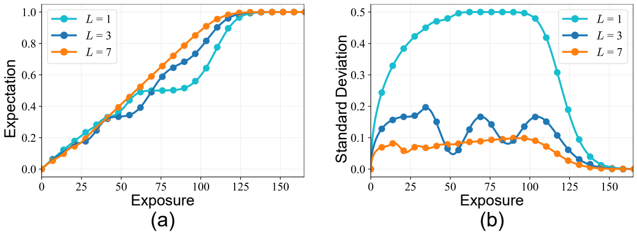

Figure 12. (a) and (b) are the curves of expectation and standard deviation, respectively. The curve plotted with theoretical analysis is represented by solid lines. The spikings generated by our simulator are marked with dots. The unit for the exposure is the number of photons per interval per pixel.

where S_i^L denotes the i-th spiking in $\{S^L\}$. The variance is driven by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\frac{S_i^L}{L} - \mu)^2. \tag{45}$$

The real-captured spikings are collected in a dark room (0LUX). The positions of all the objects and the light source remain fixed, while the intensity of the light is adjustable. We establish 30 exposure levels for the light source, and use a 1-bit spiking camera to capture the spikings that emitted within one second. In this way, we obtain 30 sets of spiking sequences with the size of $(1000 \times 1000) \times 20,000$ under varying exposure levels. Then, as shown in Fig. 11, we reconstruct 30 images using Eqn. (44). We use Eqns. (44) and (45) to calculate the expectation and variance, then use Eqns. (4) and (5) to perform curve fitting on the real-captured spikings, and the obtained results are shown in Fig. 2.

We further design a simulator for the multi-bit spiking camera to validate the characteristics of multi-bit spikings. The simulator restores the possible noise of the multi-bit spiking camera, and we simulate the process of the accumulation of photon-generated electrons and the emission of spikings. In the multi-bit spiking simulator, we set $Q_{\rm r} = 104$, $Q_{\rm b} = 1$, $Q_{\rm d} = 0.005$, and set 30 evenly distributed exposure levels. We also collect 20,000 spikings fired within 1 second. In our theoretical formulations, we set $Q_{\rm r} = 105$, $\mu_{\rm d}$ as 0.05, and set the variance of all noise distributions to 0.005. The curves under this setting and the spikings are shown in the Fig. 12. The theoretical formulations are capable of fitting the simulated spikings. In the future, we anticipate an update to the multi-bit readout mechanism, enabling us to achieve more compelling validation.

11. Additional implementation details

About the RMB spikings, we set the size of the spiking plane to be 500×500. At each readout point, we read out 4×500 three-bit spikings and 496×500 single-bit spikings. For the following readout point, the indices of the three-bit spikings are shifted downward by two lines. The encoder that is used to extract multi-scale features from $\mathbf{I}_{\rm s}(i)$ contains three blocks. Each block of the encoder contains two convolutional layers and a 2×2 maxpooling layer. Each convolutional layer is followed by batch normalization [3] and ReLU [1]. The kernel size of all the convolutional layers is 5×5. The output channels of the three blocks in the encoder are 24, 32, and 48, respectively. The encoder for extracting multi-scale features from $\mathbf{I}_{\rm m}(i)$ is nearly the same, with the exception of the input channel of the first convolutional layer. As for the cross-bit attention block, we first employ cyclic shift to obtain $(2 \times w + 1)^2$ feature pairs, and concatenate all the feature pairs. Then, we utilize two convolutional layers to learn the weight masks. The cross-time attention is simi-

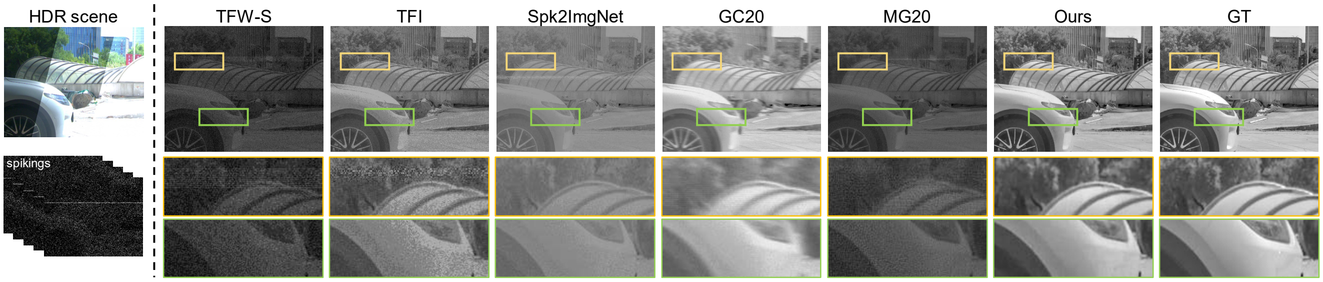

Figure 13. Additional visual equality comparison of synthetic data between the proposed method and compared methods: TFW-S [7], TFI [7], Spk2ImgNet [6], GC20 [2], and MG20 [5]. The HDR scene is captured by alternating exposures.

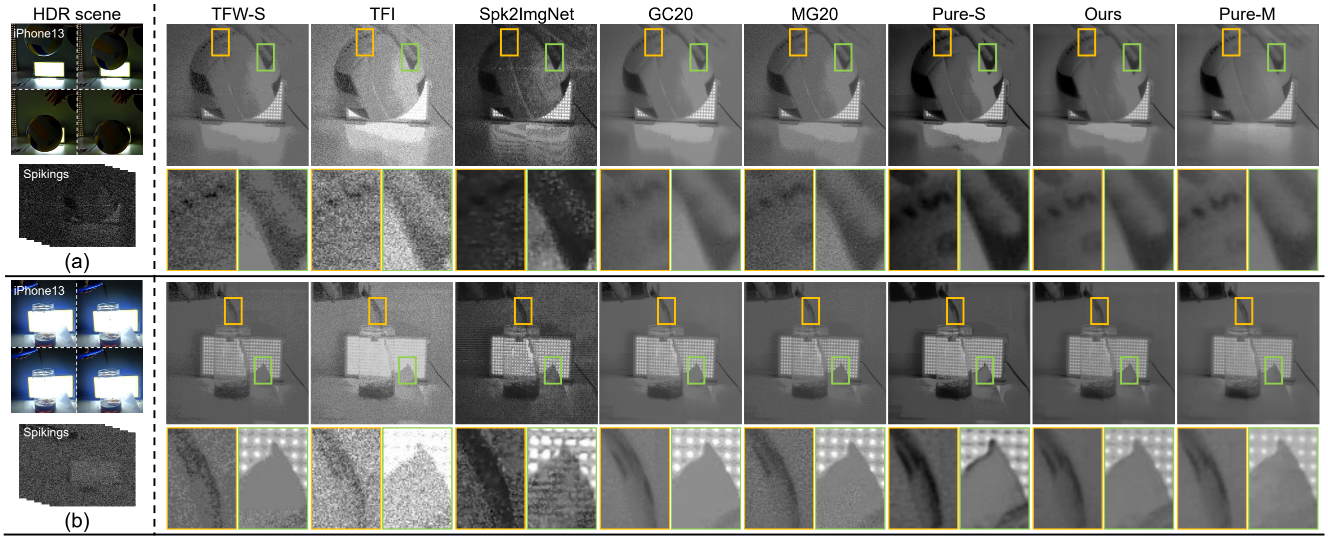

Figure 14. Additional visual equality comparison of real-synthetic data. The four frames captured by iPhone13 are used to illustrate HDR scenes.

lar to the cross-bit attention, and the difference is described in Eqn. (12). We implement our model with PyTorch, and use ADAM optimizer [4] during the training process. We adopt two NVIDIA GeForce RTX 3090 to train our model.

12. Additional qualitative results

In this section, we present additional visual comparisons on synthetic and real-synthetic data. As shown in Fig. 13, the results generated by TFW [7], TFI [7], GC20 [5] and MG20 [5] are noise-contaminated. The textures in the output generated by Spk2ImgNet [6] lack richness. Conversely, our method yields better texture details in the output. In Fig. 14 (a), we propel a volleyball towards the ground, with a bright LED array in the background. Our method outperforms the comparison methods in reconstructing the texture details on the volleyball. In Fig. 14 (b), we pour a cup of Cola into a transparent glass. Despite the deep color of the Cola, our method is still able to reconstruct the details of the flowing liquid.

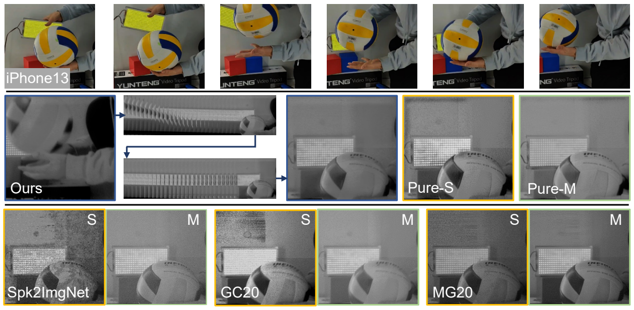

Figure 15. Results on more challenging scenes. We input pure-single-bit and pure-multi-bit spikings to the compared methods, and the results are denoted by “S” and “M” correspondingly.

In Fig. 15, we present a set of results obtained from real-synthetic data captured in the condition involving both rapid camera shake and object motions. For results display purposes, we uniformly select $\frac{1}{5}$ from the total 290 frames, which demonstrates the effectiveness of our approach in han-

Table 4. The performance of compared methods with the input of pure-single-bit and pure-multi-bit spikings.

Metric / Method	PSNR↑	SSIM↑	HDR-VDP3↑	HDR-VQM↓
TFW-S	9.12/14.48	0.411/0.592	6.241/6.803	0.932/0.852
TFW-L	13.21/15.26	0.775/0.802	7.458/7.743	0.631/0.562
TFI	17.22/——	0.693/—-	7.125//—-	0.854/—-
Spk2ImgNet	16.02/20.90	0.724/0.816	7.639/7.983	0.741/0.606
GC20	19.09/22.58	0.811/0.843	7.333/7.702	0.616/0.564
MG20	16.21/20.73	0.743/0.792	7.619/7.952	0.745/0.633
Ours	19.37/26.62	0.867/0.904	7.502/8.182	0.315/0.084

dling more challenging scenes. We additionally show the fair comparison on real-synthetic data in this figure. The results obtained from pure single-bit and pure multi-bit data are denoted by "S" and "M" correspondingly.

For more comprehensive comparison, we input pure single-bit and pure multi-bit spikings to compared methods. The quantitative evaluation is presented in Table 4. In this table, we use "/" to separate the scores corresponding to pure single-bit and pure multi-bit spikings. TFI does not support pure multi-bit input. The results demonstrating that even under identical input conditions, our approach still achieves competitive performance.

13. Limitations

It is noted that high dynamic range in high-speed scenes is a relative concept since the photons accumulated within an extremely short period are limited. Improving the dynamic range as much as possible without reducing the frame rate is worth exploring. Since the spiking camera available to us has not undergone hardware upgrades to enable the proposed RMB readout mechanism, the real data simulation is conducted through spatial and temporal aggregation. We anticipate a hardware update in the upcoming future, which will facilitate more realistic validations of our method.

References

[1] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 315–323, 2011. 4

[2] Abhiram Gnanasambandam and Stanley H Chan. HDR imaging with quanta image sensors: Theoretical limits and optimal reconstruction. *IEEE Transactions on Computational Imaging*, 6:1571–1585, 2020. 5

[3] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. of International Conference on Machine Learning*, pages 448–456, 2015. 4

[4] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. of the International Conference on Learning Representations*, 2015. 5

[5] Sizhuo Ma, Shantanu Gupta, Arin C Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Transactions on Graphics*, 39(4):79–1, 2020. 5

[6] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proc. of Computer Vision and Pattern Recognition*, pages 11996–12005, 2021. 5

[7] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proc. of Computer Vision and Pattern Recognition*, pages 1438–1446, 2020. 5