nature portfolio

Peer Review File

A stochastic encoder using point defects in two-dimensional materials

Corresponding Author: Professor Saptarshi Das

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Authors make an attempt to understand implication of defects on device reliability using atomistic imaging, density functional theory calculations, device modeling and low temperature transport experiments. They further leverage the existing point defects in 2D materials to accelerate a stochastic spiking neural network based inference engine. Please find my comments below;

Major Comments
1. Authors claim to investigate influence of defects on reliability of FET devices through random telegraph signals. To investigate the reliability, the effect of varying defect concentration on RTS should be studied. However, authors did not conduct any experiment where in the reliability of the devices were improved or degraded w.r.t defect concentration and origin. Please provide the experimental data to support your claim.

2. It is not clear whether the training and inference on Medical MNIST data set was carried out in simulation or on fabricated devices. If done on simulation, how does the simulated spiking neural devices correlate with the fabricated devices? Please provide further justification.

3. Authors need to provide references for observations made on Raman spectra.

4. Authors need to explain or provide references for the correlation of dependence of $\overline{(\tau_s)}$ and $\overline{(\tau_e)}$ on Vg and the location of point defects.

5. While realizing afferent neurons, authors argued that $V_{(BG)}$ can not be used as input and used I_{DS} as input. How was the I_{DS} varied without varying $V_{(BG)}$? How can this be implemented in real applications? Please provide explanations for the same.

Minor Comments
1. The inset temperature value in Fig. 2c needs to be corrected.
2. Axes labels in Fig. 3 e) & f) are not visible, figure needs to be enlarged. Please improve the quality of images so that readers can understand them.
3. Typing mistake needs to be corrected in “RTS dynamecs and defect correlation” section 7th line (replace “4V” by “-4V”)

Reviewer #2

(Remarks to the Author)
This is the very interesting study of exploring and utilizing defects in the 2D materials for computing, which particularly demonstrates unique advantages for noise-tolerant applications. The study highlights the potential and significance of understanding and leveraging defects in 2D materials. The study is comprehensive and solid. The manuscript is also well written and clearly presented. I have the following comments to hopefully help to improve the quality of the manuscript:

1. The labelling of the defects in Fig. 1(b) is not very clear. It is suggested to improve the quality of this figure.
2. In Fig. 2, to confirm the channel length of 20 nm, it is suggested to provide a zoom-in SEM or TEM image at the channel region.
3. In Fig. 2(c), the temperature should be 300 K instead of 15 K. Please change accordingly.
4. From Fig. 2 (c) and (d), it seems that improvement of the typical FET electrical performance, i.e. SS, transconductance, etc, is needed. Is there any relationship between these typical device performance parameters to the design of biomimetic afferent neuron?
5. Fig. 3(e) is not clear, please modify and make it clearer.
6. In Fig. 3(d), please make it clear which points are τc and which points are τe.
7. For Fig. 3(e), more details can be given regarding the TCAD simulation.
8. How much difference is the RTS pattern for devices at the different locations of the wafer?
9. It seems that the demonstrated devices for SSNN only works at cyro-temperature, i.e. 15 K, which may impact its broad applications. Please comment on this and suggest possible solutions?
10. For the SNN training and inference results, have the device-to-device variation and cycle-to-cycle variations been considered for simulation? If not, how would that impact the results?
11. It should be nice to give a short description of how the deterministic encoding was implemented.

Reviewer #3

(Remarks to the Author)
In this article, H. Ravichandran et al. reported a thorough study of defects and RTN in WSe2 FETs, and based on these devices, they also demonstrated a stochastic encoder in an inference scene. This paper is well-organized, solid, and will be of great interest to the 2D community. So, I agree to accept this paper for publication in Nature Communications. However, there are still some questions that need to be addressed:
1. It's a good idea to take advantage of the defects in 2D materials, but the authors should make it clear how many defects (neither too many nor too few) could support the following properties. I suggest the authors provide more quantitative evidence of defects, e.g., Dit (Density of interface states), and importantly, compare it to silicon devices.
2. The authors reported RTNs of WSe2 devices at low-temperature. Is it a challenge to realize RTN at room temperature? Or the defects in the device are too many?
3. Please provide vital statistics data of measured FETs, not just RTNs. For example, collect transfer curves on those with RTN data.
4. It is multi-level RTNs, not 2-state ones, that always emerge in the case of more-than-one defects. Please give a thorough analysis of the RTN results and make comments on this topic.
5. Please provide comments on the scalability of afferent neuron block in Fig.4d. How to optimize the device to save RC differentiators and comparators?
6. In the last section, "Application of stochastic encoding in biomedical imaging", the author should claim the results were from simulation, not experiments.
7. In Fig.5e and Extended Data Fig. 16, the stochastic-SNN outperformed deterministic-SNN on noise images. Could the authors provide more examples of images with different deviations, and also the corresponding results after deterministic encoding and stochastic encoding?
8. In Fig. 2c, the temperature condition was wrongly Annotated.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Dear Authors,
Thank you for providing the additional information and making the necessary changes in the manuscript.
With these updates, the goal of the study has become bit more clear. However, we still have a major comment related to the SNN training and inference accuracy of the presented SNN encoder.
One of the primary objectives of the manuscript is to demonstrate enhanced inference accuracy for noise-inflicted medical-MNIST images through a stochastic encoder. The accuracy of the inference not only depends on the presence of noise in the medical images but also on the variability of the encoder. Since the encoder in this study is characterized by the stochastic generation of spikes through WSe2 FETs, the effect of cycle-to-cycle (C2C) and device-to-device (D2D) variation on the inference accuracy is crucial.
In supplementary figure 17 you have presented the effect of Gaussian noise on inference accuracy by showing inference accuracy versus σ. However, it is unclear whether the Gaussian noise was introduced in the MNIST dataset images or in the encoding parameter $\bar{\tau}$_spike. The statistical study needs to be conducted with noise in the encoding parameter $\bar{\tau}$_spike.
If the study was indeed carried out by incorporating noise in $\bar{\tau}$_spike through D2D and C2C variation, then conducting the study over only three cycles and two devices is statistically insignificant. For a statistically significant study, σ should be calculated over at least 100 cycles and 10 devices.

We look forward to your response addressing this concern.
Best regards,

Reviewer #2

(Remarks to the Author)
The authors have addressed the questions well. I am good with publishing this version.

Reviewer #3

(Remarks to the Author)
I have no more questions for the revision. I agree to accept this article for publishing on Nature Comm.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
No more comments. Thank you for addressing my questions.

Reviewer #1 (Remarks to the Author):

Authors make an attempt to understand the implication of defects on device reliability using atomistic imaging, density functional theory calculations, device modeling and low temperature transport experiments. They further leverage the existing point defects in 2D materials to accelerate a stochastic spiking neural network-based inference engine. Please find my comments below.

We would like to thank the reviewer for their thorough reading and acknowledgment of our work. We are happy to provide further information as necessary.

Major Comments:

1. Authors claim to investigate the influence of defects on the reliability of FET devices through random telegraph signals. To investigate the reliability, the effect of varying defect concentrations on RTS should be studied. However, the authors did not conduct any experiment where in the reliability of the devices was improved or degraded w.r.t defect concentration and origin. Please provide experimental data to support your claim.

We understand the reviewer's concern and we apologize for any misunderstanding that led to the claim that this study investigates the influence of point defects of FET reliability. The study is rather geared towards harnessing the reliability concerns stemming from the presence of defects in these 2D material-based devices used for conventional computing, to implement neuromorphic computing applications. While we agree with the reviewer that we are not providing a comprehensive understanding of the impact of defects on device reliability, we appreciate the reviewer's comment and have now performed a basic experimental study on this topic, presented below. Note that the primary aim of this manuscript is to establish a correlation between RTNs n ultra-scaled WSe_2 FETs and inherent growth defects, and to explore the use of RTNs for stochastic neuromorphic computing.

To explore the effects of different defect concentrations, we introduced defects in the WSe_2 FETs through laser irradiation by adjusting the power and exposure time of a 532 nm laser typically used in Raman spectroscopy. We measured the post-irradiation RTN in five ultra-scaled WSe_2 FETs after exposure to laser powers ranging from 0.1% to 50% for an exposure time of 10s. All WSe_2 FETs have L_{CH}= 20 nm and W_{CH}= 500 nm. **Fig. R1a** shows the dual-sweep transfer characteristics obtained before and after laser irradiation at 300 K and after laser irradiation at 15 K. **Fig. R1b** shows the RTN traces in I_{DS} measured at different back-gate biases (V_{BG}) using a sampling rate of 2 ms at 15 K and **Fig. R1c** show the corresponding extracted characteristic time constants τ_c and τ_e plotted as a function of V_{BG}. The RTN traces clearly show that the spike-like RTN behavior is evident in devices exposed to laser powers of 0.1% and 1%, with τ_c and τ_e being independent of gate bias, similar to as-fabricated devices. This indicated that these levels of laser power do not significantly alter the inherent defect density and the defect dynamics of inherent defects in these WSe_2 FETs. Conversely, devices exposed to 25% and 50% laser powers exhibited more rectangular

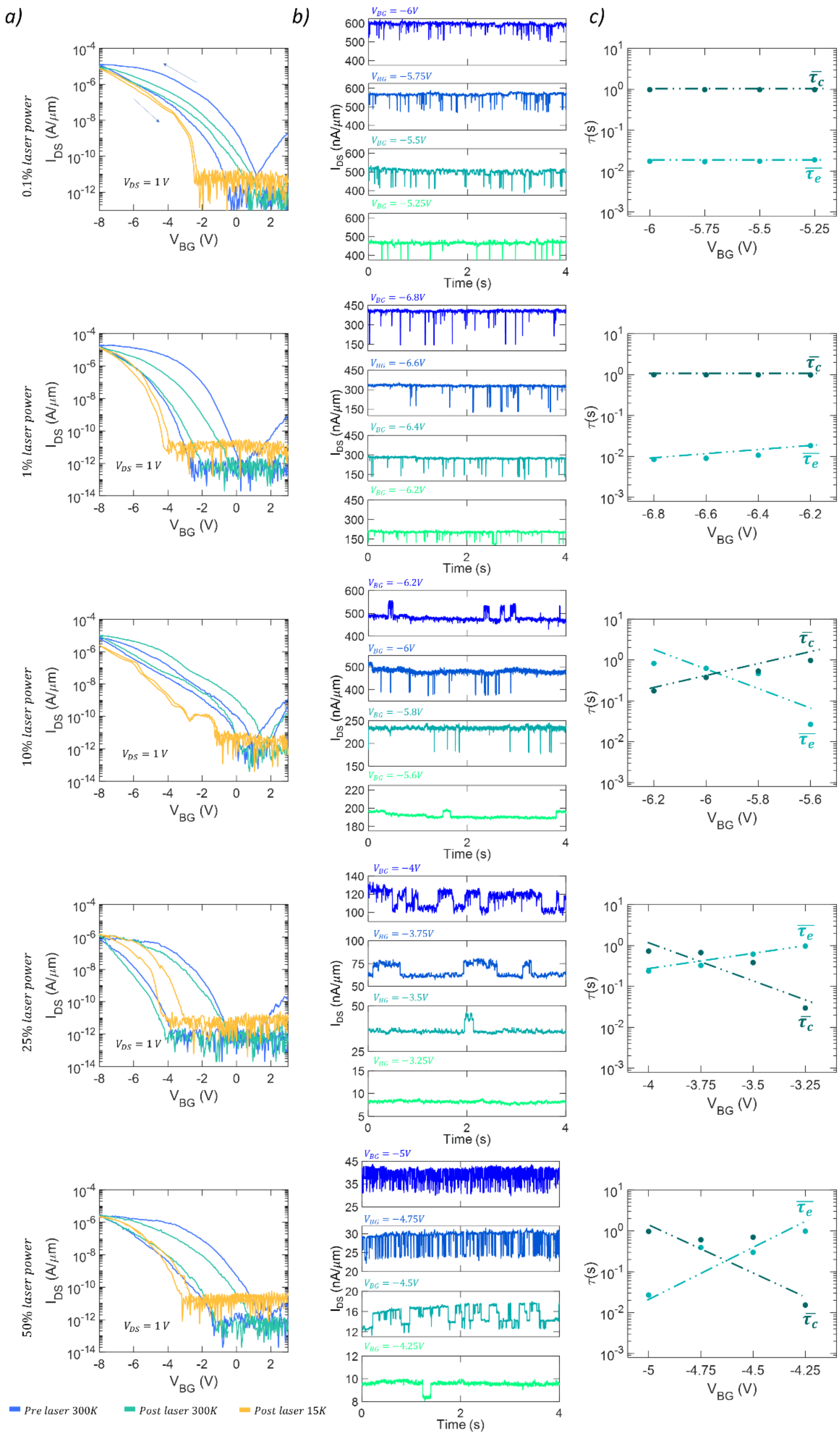
a)
b)
c)
0.1% laser power
1% laser power
10% laser power
25% laser power
50% laser power
I_{DS} (A/μm)
V_{BG} (V)
$V_{DS} = 1\,V$
I_{DS} (nA/μm)
Time (s)
$V_{BG} = -6V$
$V_{BG} = -5.75V$
$V_{BG} = -5.5V$
$V_{BG} = -5.25V$
$V_{BG} = -6.8V$
$V_{BG} = -6.6V$
$V_{BG} = -6.4V$
$V_{BG} = -6.2V$
$V_{BG} = -6.2V$
$V_{BG} = -6V$
$V_{BG} = -5.8V$
$V_{BG} = -5.6V$
$V_{BG} = -4V$
$V_{BG} = -3.75V$
$V_{BG} = -3.5V$
$V_{BG} = -3.25V$
$V_{BG} = -5V$
$V_{BG} = -4.75V$
$V_{BG} = -4.5V$
$V_{BG} = -4.25V$
τ(s)
$\overline{\tau_c}$
$\overline{\tau_e}$
Pre laser 300K
Post laser 300K
Post laser 15K

Figure R1. Impact of laser irradiation on device characteristics and defect dynamics in monolayer WSe_2 FETs. a) *Transfer characteristics of five ultra scaled WSe_2 FETs with L_{CH}=20 nm and W_{CH}=500 nm each measured at three different conditions: at 300 K before laser irradiation, at 300 K post laser irradiation and at 15 K post laser irradiation. The laser irradiation was performed by exposing the channel of the devices to a 532 nm laser for a duration of 10 s using a Raman spectroscopy setup for a filtered laser power of 0.1%, 1%, 10%, 25% and 50% at T=300K.* ***b)*** *RTN traces obtained from the five WSe_2 FETs post laser irradiation of corresponding power measured at T=15K. The V_{BG} used for obtaining the RTN traces are provided in the inset.* ***c)*** *$\overline{\tau_e}$ and $\overline{\tau_c}$ as a function of V_{BG} obtained through exponential fits to the experimental data obtained for each of the five WSe_2 FETs shown in* ***b)****. It is evident that at low laser powers of 0.1% and 1%, the characteristic time constants of the defects τ_c and τ_e are independent of the applied V_{BG} suggesting that low laser power does not alter the inherent dynamics of growth defects present in WSe_2 which leads to the observation of gate bias independent RTN. At laser powers >25%, the characteristic time constants show a V_{BG} dependence, indicating that the high-power laser exposure has led to the formation of oxide traps which leads to the observation of gate bias dependent RTN. This also points to the fact that high-laser power exposure alters the inherent defect dynamics observed in scaled WSe_2 FETs.*

RTN traces, with τ_c and τ_e showing dependence on V_{BG}. This suggests that the RTNs observed in these cases originate from defects in the dielectric rather than the channel. In other words, higher magnitude of laser powers induces oxide defects, which modify the dynamics of the observed RTN. We have also extracted the density of interface traps (D_{IT}) for these WSe_2 FETs. The D_{IT} is extracted from the subthreshold slope (SS) of the device transfer characteristics. The SS is given by the equation,

$$SS = m\frac{k_B T}{q} ln10 \left(\frac{mV}{decade}\right) \quad [1]$$

$$m = \left(1 + \frac{C_S + C_{IT}}{C_{ox}}\right) \quad [2]; \; C_S \approx 0 \text{ for fully depleted ultra-thin body 2D FETs}$$

$$C_{IT} = qD_{IT} \quad [3]$$

Where k_B is the Boltzmann's constant, T is the temperature in K, m is the body factor, C_{IT} is the capacitance due to the interface traps, C_{ox} is the oxide capacitance and D_{IT}, the density of interface traps. From the SS, C_{IT} can be calculated and the D_{IT} can be determined using **Eq.3**. Post laser-irradiation with laser powers >10%, D_{IT} value increased from 2×10^{13} $eV^{-1}cm^{-2}$ to 4.2×10^{13} $eV^{-1}cm^{-2}$, a 2x increase as shown in **Fig. R2**, which further points to the increased defect density due to the formation of oxide traps that leads to the observed RTN. While it is necessary to study the impact of changes in defect concentration on device reliability, a more thorough study on this topic is beyond the scope of this work and we are happy to address this concern in a future study.

2. It is not clear whether the training and inference on the Medical MNIST data set was carried out in simulation or on fabricated devices. If done on simulation, how do the simulated spiking neural devices correlate with the fabricated devices? Please provide further justification.

We apologize for any confusion caused by the omission of the details of the simulation results. To rectify this, we have thoroughly updated the manuscript to provide a clearer understanding of our methodologies. In this study, we have employed a spiking neural network (SNN) to categorize the medical MNIST dataset, which comprises six unique classes of medical images including abdominal, breast, chest CT, chest X-ray, hand X-ray, and head CT, as depicted in **Fig. R3a**. The

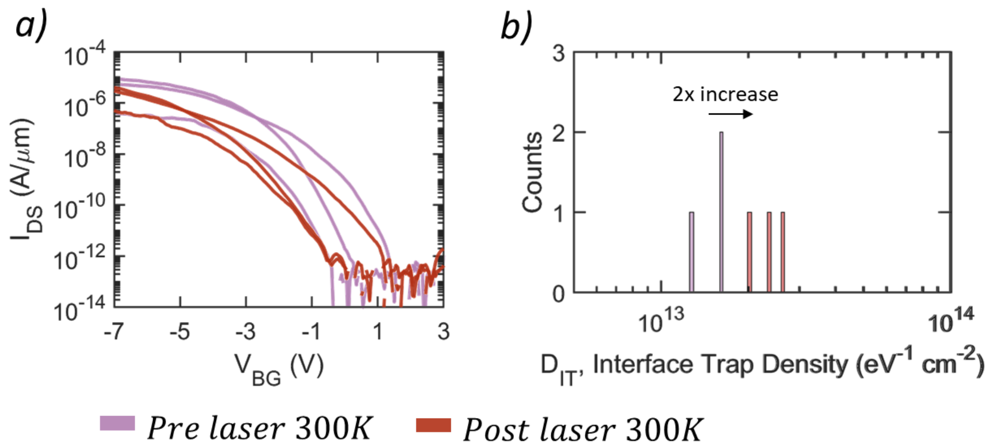

Figure R2. *Impact of laser irradiation on the density of interface traps.* ***a)*** *Transfer characteristics of three ultra scaled WSe_2 FETs with L_{CH}=20 nm and W_{CH}=500 nm each measured at 300 K before and after laser irradiation using a 532nm laser for a duration of 10s at a filtered laser power of 10%, 25% and 50% at T=300K.* ***b)*** *Plot showing the density of interface traps (D_{IT}) measured pre and post laser irradiation. The measured D_{IT} values increased from 2×10^{13} $eV^{-1}cm^{-2}$ to 4.2×10^{13} $eV^{-1}cm^{-2}$, a 2x increase, pointing to the increased defect density due to the formation of oxide traps.*

dataset contains 48,000 images, each with a resolution of 64×64 pixels. Since SNNs require spike trains as inputs, all pixel intensities are converted into spike trains where the mean time between spikes ($\bar{\tau}_{spike}$) is inversely related to pixel brightness; brighter pixels result in shorter times between spikes, and darker ones longer. **Fig. R3b** illustrates the linear relationship between the drain current (I_{DS}) and $\bar{\tau}_{spike}$. We have formulated an empirical model using **Eq.4** to replicate $\bar{\tau}_{spike}$, with I_{DS} linking our experimental and simulation work.

$$\bar{\tau}_{spike} = \tau_0 \times exp\left(-\frac{I_0}{I_{DS}}\right) \qquad [4]$$

where I_0 and τ_0 are the fitting parameters. Given the extensive task of generating spike trains for the entire set of 64 × 64 × 48000-pixel intensities across all images, we used **Eq.4** to encode this data as shown in **Fig. R2c**. By integrating empirical modeling with simulation, we connected the fabricated devices to a stochastic encoder for the SNN for classifying medical images, which are often subject to external noise.

3. Authors need to provide references for observations made on Raman spectra.

We thank the reviewer for bringing up this point. We have now added the following references in the main manuscript to support our observation obtained from the low-temperature Raman spectroscopy [1-3].

[1] Y.-C. Lin et al., "Realizing Large-Scale, Electronic-Grade Two-Dimensional Semiconductors," ACS Nano, vol. 12, no. 2, pp. 965-975, 2018/02/27 2018, doi: 10.1021/acsnano.7b07059.

[2] S. Lippert et al., "Influence of the substrate material on the optical properties of tungsten diselenide monolayers," 2D Materials, vol. 4, no. 2, p. 025045, 2017/03/09 2017, doi: 10.1088/2053-1583/aa5b21.

[3] Z. Wu et al., "Defect Activated Photoluminescence in WSe2 Monolayer," The Journal of Physical Chemistry C, vol. 121, no. 22, pp. 12294-12299, 2017/06/08 2017, doi: 10.1021/acs.jpcc.7b03585.

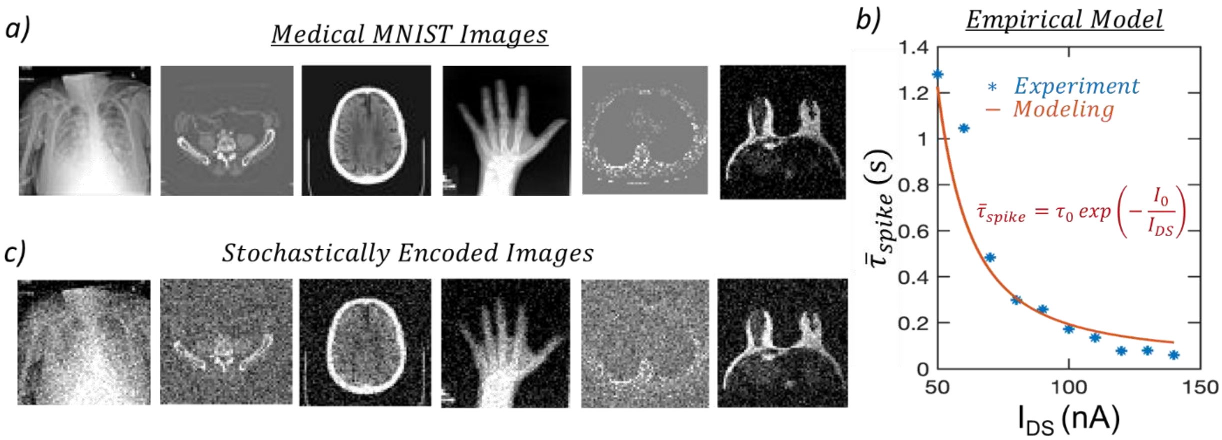

Figure R3. Stochastic encoding of Medical MNIST images using empirical modeling. a) *Representative image from each class in the medical-MNIST dataset corresponding to abdomen, breast, chest CT, chest X-ray, hand X-ray, and head CT. Medical-MNIST contains 48,000 images, with each classes containing 8,000 images of size 64 × 64 pixels.* ***b)*** *Plot of $\bar{\tau}_{spike}$ as a function of I_{DS} obtained experimentally along with an empirical model replicating the relationship between $\bar{\tau}_{spike}$ and I_{DS} obtained experimentally which serves as the basis for encoding the medical MNIST images used for the stochastic SNN based inference acceleration.* ***c)*** *Stochastically encoded representative medical MNIST images from each class utilizing the empirical model.*

4. Authors need to explain or provide references for the correlation of dependence of (τ_s)¯and (τ_e)¯on Vg and the location of point defects.

We would like to thank the reviewer for bringing this to our attention. We have now performed TCAD modeling to provide insight on the location of the defect and the observed gate bias independence of the capture and emission time constants. We have added the following discussion to the main manuscript.

The hypothesis that the dominant defects causing the observed RTN originate within the WSe_2 channel is further supported by TCAD modeling. This modeling shows a qualitatively similar trend of gate bias-independent capture and emission time constants for charge traps within the WSe_2 channel. Our modeling analysis further indicates that the defects associated with these charge traps possess trap energy (E_T) of approximately 100 meV and are located above the valence band maximum (E_V). Additionally, these charge traps have a relaxation energy (E_{relax}) smaller than 150 meV. It should be noted that these charge transition levels and relaxation energies agree well with the values calculated for hole trapping at Se_W and V_W using DFT calculations as shown in **Supplementary Figure 2** in the revised manuscript. Additionally, the TCAD modeling also reveals that the defect is to be located near the drain contact within the channel. These findings provide robust evidence that the observed RTNs result from defects present in the WSe_2 channel.

We have also added the following reference [4] that throws light on the gate bias independence of the characteristic time constants of channel defects in the revised manuscript.

[4] B. Stampfer et al., "Characterization of Single Defects in Ultrascaled MoS2 Field-Effect Transistors," ACS Nano, vol. 12, no. 6, pp. 5368-5375, 2018/06/26 2018 doi: 10.1021/acsnano.8b00268.

5. While realizing afferent neurons, authors argued that V_(BG) cannot be used as input and used I_DS as input. How was the I_DS varied without varying V_(BG)? How can this be implemented in real applications? Please provide explanations for the same.

First, we would like to point out that, V_{BG} could potentially be used as an input so long as the characteristic capture and emission time constants follow a monotonic behavior with respect to V_{BG}. Next, we would like to apologize for the lack of clarity in explaining the experimental setup used for obtaining the voltage spikes using I_{DS} as the input. Please note that V_{BG} is kept constant while the device is provided an input current I_{DS} from the parameter analyzer used for the measurements to obtain the voltage spikes at the drain terminal. This is similar to using a current source to measure the voltage drop across the device instead of using a voltage source to measure the current through the device. The voltage spikes obtained from the device are then passed through the peripheral spike digitization circuit consisting of a differentiator and a comparator to convert the analog output voltage spikes into digital spike trains for practical applications. We have now added further details on how the voltage spikes were obtained in the revised manuscript.

Minor Comments:
1. The inset temperature value in Fig. 2c needs to be corrected.
We apologize for the incorrect labels in Fig. 2c. We have corrected the insert values in Fig. 2c in the revised manuscript.

2. Axes labels in Fig. 3 e) & f) are not visible, figure needs to be enlarged. Please improve the quality of images so that readers can understand them.

We apologize for the mislabeling of figure labels and the low figure quality. We have made changes to the labels in Fig. 3e and Fig. 3f and have improved the quality of the figures for better readability.

3. Typing mistake needs to be corrected in “RTS dynamics and defect correlation” section 7th line (replace “4V” by “-4V”)

We are sorry for the typo. We have made the necessary correction in the revised manuscript

Reviewer #2 (Remarks to the Author):

This is a very interesting study of exploring and utilizing defects in 2D materials for computing, which particularly demonstrates unique advantages for noise-tolerant applications. The study highlights the potential and significance of understanding and leveraging defects in 2D materials. The study is comprehensive and solid. The manuscript is also well written and clearly presented. I have the following comments to hopefully help to improve the quality of the manuscript:

We would like to thank the reviewer for the appreciation of our work and for recommending our work for publication. We are happy to provide further information as necessary.

1. The labelling of the defects in Fig. 1(b) is not very clear. It is suggested to improve the quality of this figure.

We thank the reviewer for this suggestion. We have now replaced Fig. 1(b) in the main manuscript with a better-quality image and defect labeling. We have also added figures 1(b) showing zoomed-in STEM image corresponding to each defect. The zoomed-in STEM images are shown in **Fig. R4**.

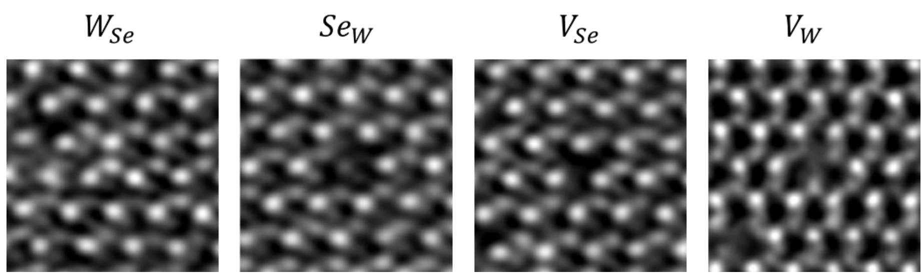

Figure R4. *Zoomed-in STEM image revealing the presence of several points defects such as V_{Se}, Se_W, W_{Se} and V_W.*

2. In Fig. 2, to confirm the channel length of 20 nm, it is suggested to provide a zoom-in SEM or TEM image at the channel region.

We thank the reviewer for this suggestion. We have now added a zoomed in version of the SEM showing the channel region in **Supplementary Figure 4**. The channel length turned out to be ~ 19.8 ± 0.05 nm. The zoomed-in SEM image is depicted in **Fig. R5**.

3. In Fig. 2(c), the temperature should be 300 K instead of 15 K. Please change accordingly.

We apologize for overlooking these formatting errors. We have now made the necessary corrections to the main manuscript.

4. From Fig. 2 (c) and (d), it seems that improvement of the typical FET electrical performance, i.e. SS, transconductance, etc, is needed. Is there any relationship between these typical device performance parameters to the design of biomimetic afferent neuron?

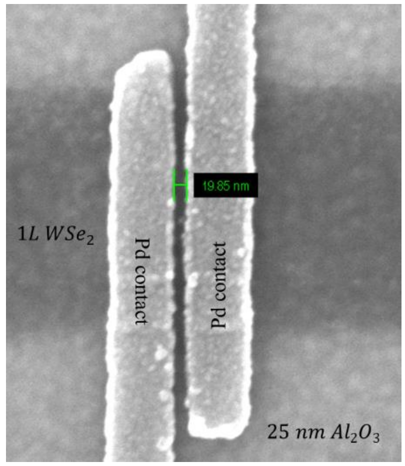

Figure R5. *Zoomed-in SEM image revealing an* L_{CH}*=19.8* $\pm$ *0.05 nm.*

We agree with the reviewer that the performance of the WSe_2 FET characteristics needs improvement when aiming for competitive performance for applications in digital logic. We would like to point out that the devices used in this study exhibits good electrical performance with median *ON* currents reaching > $1\mu A/\mu m$ with a current on/off ratio >10^5. We have also been able to achieve *ON* currents as high as $40\mu A/\mu m$ in p-type WSe_2 FETs, the results of which are summarized in our earlier study [5]. However, we believe that there is room for further improvement through optimization of growth, contact engineering, doping strategies and other techniques. Nevertheless, the ultra-scaled WSe_2 FETs are representative of devices in existing demonstrations. Furthermore, at 15K, the device performance degrades due to the Schottky barrier nature.

The design of the afferent neurons involves harnessing single defects present in scaled 2D FETs, thereby the suggested operation mode as afferent neurons benefits from the high defect densities observed in our FETs. The impact of single defects follows an exponential distribution with the single defect regime typically consisting of 10 active defects that are in the tail of the distribution. These defects lead to the observation of RTN which is essential for the design of the biomimetic afferent neuron. When a large number of defects are present, the influence of a single active defect on the device characteristics that lead to RTN cannot be observed. This not only affects the device reliability adversely, but also hinders the realization of afferent neurons using these devices. The presence of no defects also prevents the construction of the afferent neuron as the defect dynamics of single defects is central to this application. Hence it is necessary to ensure that the devices utilized for afferent neurons operate in the single defect regime.

5. Fig. 3(e) is not clear, please modify and make it clearer.

We thank the reviewer for this suggestion. We have made the necessary changes to the revised manuscript.

6. In Fig. 3(d), please make it clear which points are τc and which points are τe.

We thank the reviewer for this suggestion. We have made the necessary changes to Fig. 3d in the revised manuscript.

7. For Fig. 3(e), more details can be given regarding the TCAD simulation.

We thank the reviewer for this suggestion. We apologize for the misunderstanding, but it should be noted that Fig. 3(e) shows the time constants detected in one RTN trace, as extracted using the Canny edge detection algorithm and an evaluation of the exponential distributions of the time constants. Subsequently, we compared the gate bias dependence qualitatively with TCAD modeling using a non-radiative multi-phonon model to shed light on the gate-bias independent capture and emission time constants for channel-related RTN defects in the on-state of the devices. We have included further details on the TCAD modeling in the Methods section of the revised manuscript.

8. How much difference is the RTS pattern for devices at the different locations of the wafer?

The reviewer has raised an important question. To understand the RTN patterns obtained across the wafer, we measured multiple devices that are located physically in different parts of the wafer. The findings are tabulated in **Supplementary Figure 10-11**, where we found that RTN is observed in all these devices irrespective of the location in the wafer suggesting that the plausible defects responsible for the observation of RTN is uniformly distributed across the entire wafer. Additionally, we have also utilized the RTN obtained from two different devices for stochastic spike encoding of Medical MNIST images followed by the inference of noise-afflicted Medical MNIST images using SNN as shown in **Fig R6c-d**. The inference accuracy is not affected by the device-to-device variations in the RTN patterns. While the variations in the RTN patterns obtained from different devices are intrinsic and unavoidable, the functioning of the stochastic encoder remains agnostic to it as long as the exponential dependence between $\overline{\tau}_{\text{spike}}$ and I_{DS} is maintained, which is a direct consequence of the defects that are being probed using electrical measurements, which in our case are the fairly uniformly distributed growth defects in WSe_2.

9. It seems that the demonstrated devices for SSNN only work at cyro-temperature, i.e. 15 K, which may impact its broad applications. Please comment on this and suggest possible solutions.

The reviewer has raised an excellent point. Although we have significantly reduced the channel lengths of our WSe_2 FETs to 20 nm, the channel width remains at approximately 500 nm. This larger channel width prevents operation in the single-defect regime at room temperature, as multiple defects are energetically available for random telegraph noise (RTN) generation. However, at reduced temperatures, most defects become inactive, enabling single-defect regime operation. To achieve room temperature RTN with fewer channel defects, reducing the channel width to below 50 nm would be effective. It is also worth noting that at 300K, RTN in nanoscale devices is predominantly influenced by oxide traps with time constants ranging from a few milliseconds to seconds. Channel defects only contribute to RTN if the observation time is in the pico to nanosecond range. Therefore, based on the application need, it is possible to harness either the oxide defects or the channel defects for enabling room-temperature operation of the stochastic encoder as long as both the channel length and the width are sufficiently scaled. These insights have been incorporated into the revised manuscript.

10. For the SNN training and inference results, have the device-to-device variation and cycle-to-cycle variations been considered for simulation? If not, how would that impact the results?

We thank the reviewer for raising a valid point. Following the reviewer's suggestion we have now evaluated the inference accuracy of SNN considering both device-to-device (D2D) variation and cycle-to-cycle variation (C2C). **Fig. R6a** and **Fig. R6c**, respectively, shows the C2C & D2D variations in mean inter spike interval ($\bar{\tau}_{spike}$) for various drain currents (I_{DS}) over 3 cycles and 2 different devices correspondingly. To quantify the relationship between $\bar{\tau}_{spike}$ and I_{DS}, we developed an empirical model expressed as,

$$\bar{\tau}_{spike} = \tau_0 \times exp\left(-\frac{I_0}{I_{DS}}\right)$$

where I_0 and τ_0 are fitting parameters. Subsequently, we generated spike trains using the empirical model to represent images from the Medical-MNIST dataset (M-MNIST). These spike trains were then fed into a fully trained SNN. Their classification accuracy was assessed considering both C2C and D2D variation, as shown in **Fig. R6b** and **Fig. R6d**, respectively. These results have been incorporated in **Supplementary Figure 17** in the revised manuscript. Our SNN model demonstrated consistently high accuracy in identifying the M-MNIST dataset images, even when different levels of gaussian noise with standard deviations $\sigma = 0$ to $\sigma = 0.6$ were introduced to the input image. Notably, our results indicate that the performance of our model remains unaffected by the C2C and D2D variations. The intrinsic noise tolerance of SSNNs may be responsible for this stability, underscoring the potential of SSNNs for deployment in real-world scenarios where variability is a given.

11. It would be nice to give a short description of how the deterministic encoding was implemented.

We are sorry for not providing a description of deterministic spike encoding. In deterministic spike encoding, each pixel intensity of the image is translated directly into a spike train, where the timing between each spike is precisely determined by the pixel intensity. This method employs a linear mapping of the intensity values to the interspike interval ensuring that higher intensities result in the spikes occurring more frequently. In other words, more spikes occur for higher intensity values within a predefined time window. This eliminates any variability in the spike interval or number of spikes that represents a specific intensity value.

In stochastic spike encoding, a pixel with higher intensity will, on average, generate spikes at a higher rate than a pixel with lower intensity, but the precise moments at which these spikes occur can vary. This stochasticity introduces a level of randomness that can mimic the natural variability observed in biological neural systems. We have included this discussion in the revised manuscript for readers clarification. We hope this clarification addresses the request for a description of the deterministic encoding method and further elucidates the distinctions between the deterministic and rate-based encoding strategies employed in our study.

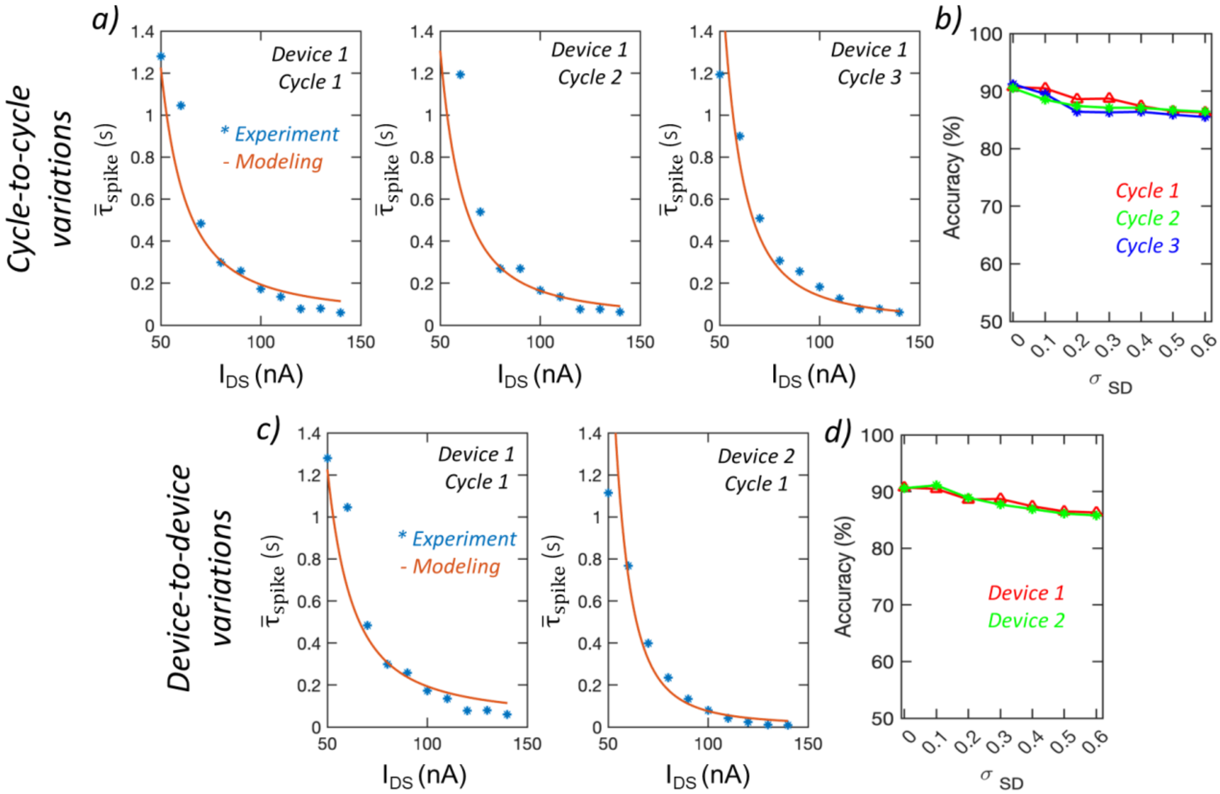

Figure R6. Impact of cycle-to-cycle and device-to-device variation on inference accuracy of medical MNIST images. a) *$\bar{\tau}_{spike}$ as a function of the input I_{DS} for one device over 3 cycles along with the fitting based on the empirical model.* ***b)*** *Corresponding inference accuracy as a function of σ for stochastic encoding. It is evident that the accuracy remains consistent when device-to-device variations are considered.* ***c)*** *$\bar{\tau}_{spike}$ as a function of the input I_{DS} for two devices for one cycle along with the fitting based on the empirical model.* ***d)*** *Corresponding inference accuracy as a function of σ for stochastic encoding. The accuracy remains agnostic to device-to-device variations indicating the robustness of the stochastic SNN for noise-tolerant inference applications.*

Reviewer #3 (Remarks to the Author):

In this article, H. Ravichandran et al. reported a thorough study of defects and RTN in WSe2 FETs, and based on these devices, they also demonstrated a stochastic encoder in an inference scene. This paper is well-organized, solid, and will be of great interest to the 2D community. So, I agree to accept this paper for publication in Nature Communications. However, there are still some questions that need to be addressed:

We would like to thank the reviewer for their appreciation of our work. We are happy to provide further details as necessary.

1. It's a good idea to take advantage of the defects in 2D materials, but the authors should make it clear how many defects (neither too many nor too few) could support the following properties. I suggest the authors provide more quantitative evidence of defects, e.g., Dit (Density of interface states), and importantly, compare it to silicon devices. (TK)

The reviewer has an excellent point. It is important to quantify the number of defects that would support the implementation of a stochastic encoder using ultra-scaled WSe_2 FETs. To this end, we have now extracted the D_{IT} from the transfer characteristics for the devices that showed RTN at low temperatures as shown in **Fig. R7**. The mean value of D_{IT} was found to be 2×10^{13} $eV^{-1}cm^{-2}$. This corresponds to about ~2000 defects in an active device area of 0.01 μm^2. This value is considerably large compared to the single defect regime where typically less than 10 active defects are present in the channel area, necessitating low temperature measurements to study the defect induced RTN dynamics. At low temperatures, such as 15K, used in our experiments, we are able to probe the channel defects as most of the oxide defects tend to freeze out at this temperature [6]. The D_{IT} values obtained from our devices are considerably higher compared to the state-of-the-art Si devices where the D_{IT} values are in the order of 1-10 × 10^9 $eV^{-1}cm^{-2}$ [7] This further justifies the fact that RTNs were not observed in these devices at room temperature. With improvements in the growth techniques for synthetic large area 2D materials, the D_{IT} of scaled 2D FETs will see a substantial reduction improving device performance and reliability. Hence, since it is necessary to ensure that the devices utilized for afferent neurons operate in the single defect regime, we had to cool them down to 15K. However, this should only be considered a temporary measure until smaller devices or devices with lower defect densities become available.

2. The authors reported RTNs of WSe_2 devices at low-temperature. Is it a challenge to realize RTN at room temperature? Or the defects in the device are too many? (TK)

The reviewer has raised an excellent point. As we have pointed out in the revised manuscript, the use of devices with comparatively large channel width necessitated the use of low temperatures for RTN measurements. It has been shown that for defect densities of the order of $10^{12}/cm^2$ for MoS_2/Al_2O_3, there could be as high as 20,000 defects in an active device area of 2.5 μm^2 [8] while the single defect limit of 10 defects can be achieved in MoS_2/SiO_2 devices by reducing the channel area to $10^{-3}\mu m^2$ [4]. This reduction in channel area enabled the study of RTN at room temperature in ultra scaled MoS_2 FETs. With the extracted D_{IT} in our devices being on the order of 10^{13}-$10^{14}/cm^2$, to observe RTN at room temperature, the channel area would have to be <$10^{-3}\mu m^2$ which would involve further scaling down the width of our devices to <50 nm. While this deals

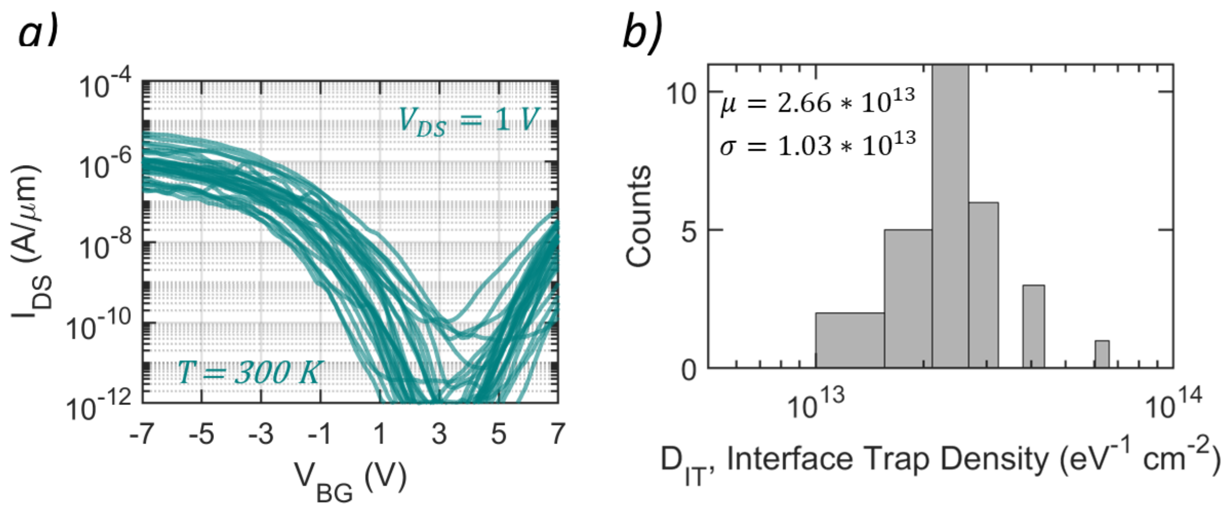

Figure R7. *Device statistics obtained from 30 WSe_2 FETs showing* ***a)*** *the transfer characteristics and* ***b)*** *the corresponding distributions for interface trap density, respectively.*

with the observation of RTNs stemming from the presence of defects in the channel, room-temperature RTN can be observed in ultra-scaled 2D FETs due to the presence of oxide-traps. In the present study we have limited our scope to harnessing RTN observed in WSe_2 FETs due to inherent growth defects and the oxide defects tends to freeze out at low temperature. With the large device dimensions used in our study it is highly plausible that room temperature measurements will not yield RTSs caused due to the presence of a large number of oxide traps.

3. Please provide vital statistics data of measured FETs, not just RTNs. For example, collect transfer curves on those with RTN data.

We apologize for skipping over the details of the FET parameters. In our revised manuscript, we have included the statistical distribution of the vital device parameters such as subthreshold slope, mobility, ON current, and threshold voltage, observed for short-channel WSe_2 FETs at room temperature. **Fig. R8a-f** depicts the distribution of the device parameters obtained from 30 ultra-scaled WSe_2 FETs. We would also like to refer the reviewer to the relevant publication that describes the FET performance metrics [5]. Additionally, we have also included the transfer characteristics of 10 WSe_2 FETs that yielded RTN shown in **Supplementary Figure 12** measured at 300 K and 15 K respectively, for a $V_{DS} = 1V$. **Fig. R9** shows the transfer characteristics of these devices.

4. It is multi-level RTNs, not 2-state ones, that always emerge in the case of more-than-one defects. Please give a thorough analysis of the RTN results and make comments on this topic. (TK)

The reviewer has an excellent point. While we agree with the reviewer that multi-level RTN is typically observed in case of the presence of more than one defect, we have limited our study to probing a single dominant defect which is most active at low temperature. The low temperature measurements allow for accessing the impact of a single dominant defect in the device while the

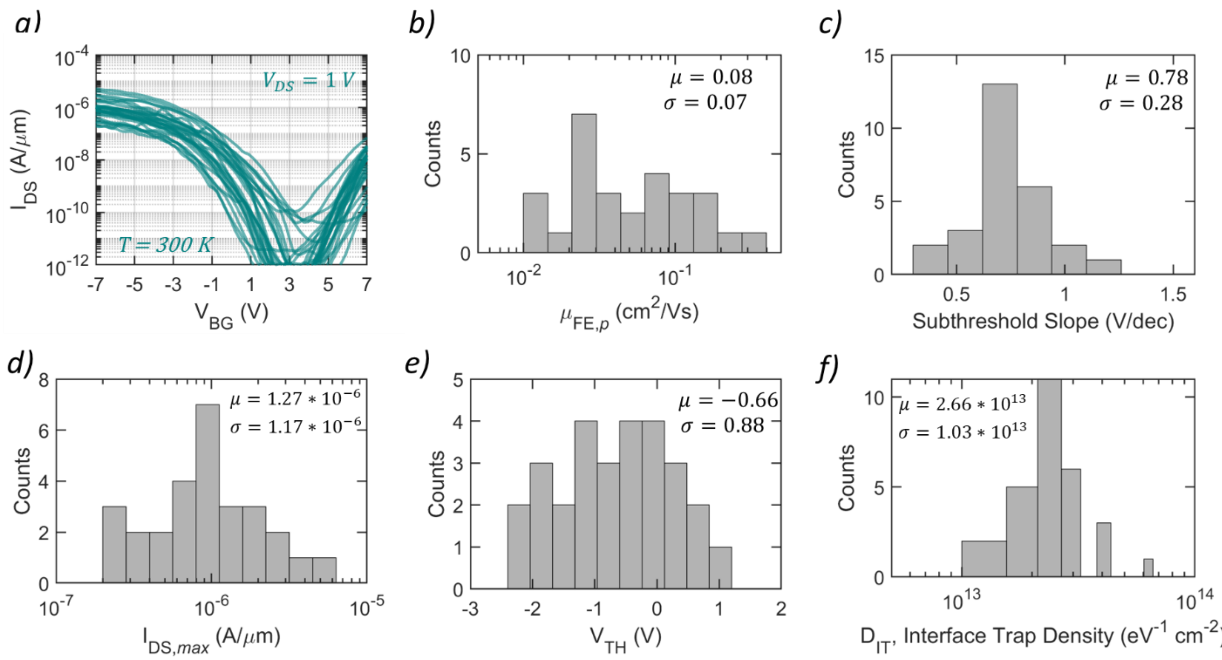

Figure R8. *Device statistics obtained from 30 WSe_2 FETs showing a) the transfer characteristics and corresponding distributions for b) field-effect mobility ($\mu_{FE,p}$), c) subthreshold slope, d) maximum ON current, e) threshold voltage and f) interface trap density, respectively.*

oxide traps are mostly frozen out [6]. In our detailed analysis of the RTN, we have considered the defect dynamics owing to the presence of a single dominant defect while ignoring the presence of multiple defects as well as other complications arising from multi-state defects. While we acknowledge that it is important to study multi-level RTN to gain a complete understanding of the defect dynamics in large-area WSe_2 FETs, we believe that this intensive study is beyond the scope of this work. In this study we wanted to provide a correlation between the observed RTN, and the inherent growth defects present in synthetic WSe_2 and to utilize the stochasticity in the device characteristics stemming from these defects for neuromorphic application. We are happy to pursue a thorough study on the defect dynamics of WSe_2 FETs in future work.

5. Please provide comments on the scalability of afferent neuron block in Fig.4d. How to optimize the device to save RC differentiators and comparators?

The reviewer has raised an important question regarding the scalability of the afferent neuron block. We have utilized the digitization circuit consisting of a differentiator followed by a comparator to ensure a constant output voltage level in the output V_{DS} while allowing $\bar{\tau}_{spike}$ to vary with the strength of the input stimulus for practical applications. While we agree that the use of these peripherals adds up to the energy and area overhead of our stochastic encoder, we believe that the digitization unit is an integral part of the afferent neuron block to enable the proper functioning of the afferent neuron. In the present demonstration, we have used external peripherals based on Si-CMOS circuits for spike digitization. For better scalability, all the peripherals could be on-chip, fabricated from 2D materials to enable the seamless operation of the stochastic encoder

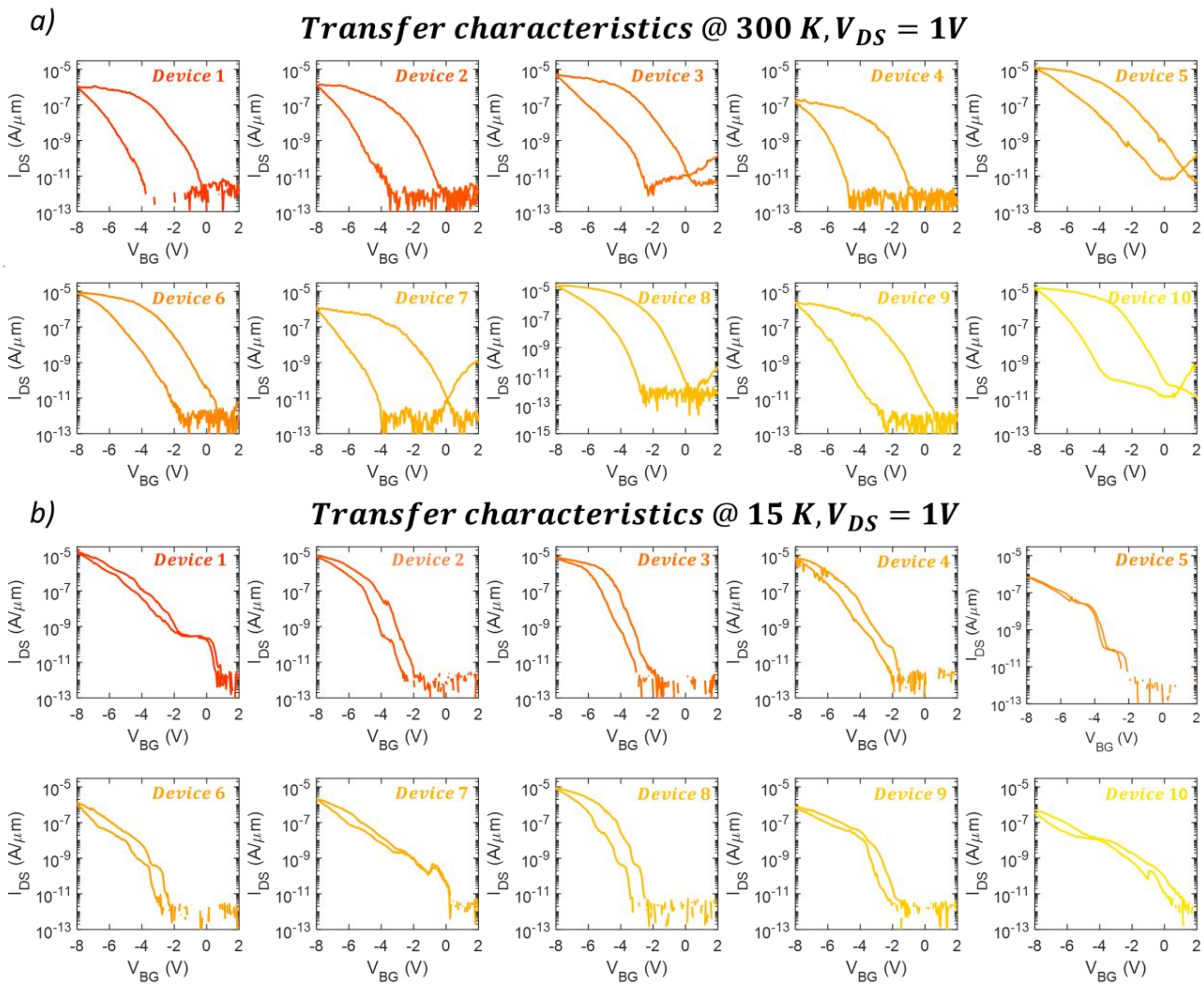

Figure R9. Transfer characteristics of multiple devices yielding RTN. *Transfer characteristics of 10 ultra-scaled WSe_2 FETs having a channel area (A_{ch}) of 0.01. μm^2 measured at **a)** 300 K and **b)** 15K, respectively, for a $V_{DS} = 1V$.*

[9, 10]. We hope to develop a complete 2D on-chip acceleration of stochastic SNN harnessing the defect dynamics in ultra-scaled 2D devices and publish the results in a future manuscript.

6. In the last section, "Application of stochastic encoding in biomedical imaging", the author should claim the results were from simulation, not experiments.

We would like to thank the reviewer for this suggestion. We have now added the details of the simulation performed for the stochastic encoding in biomedical imaging.

7. In Fig.5e and Extended Data Fig. 16, the stochastic-SNN outperformed deterministic-SNN on noise images. Could the authors provide more examples of images with different deviations, and also the corresponding results after deterministic encoding and stochastic encoding?

We appreciate the reviewer's insightful suggestion to provide more examples of images with different deviations, along with corresponding results after deterministic and stochastic encoding. In response to this suggestion, we have added new results using the widely recognized MNIST handwritten digit dataset. We chose this dataset because it allows for a clear demonstration of how our encoding methods perform with standardized, well-understood images. As requested, we encoded these images using both deterministic and stochastic methods across a range of standard deviations, illustrating the robustness of the stochastic-SNN approach under various noise conditions. These new results are shown in **Fig. R10** and have been incorporated into **Supplementary Figure 14** of the revised manuscript. Our analysis clearly indicates that the stochastic-encoded images maintain higher robustness to noise. This enhanced robustness contributes significantly to the improved classification accuracy observed with stochastic-SNN compared to deterministic-SNN. These findings further substantiate our original results and conclusions presented in **Fig. 5e** and **Supplementary Figure 13** in the revised manuscript. We believe that these additions address the reviewer's concerns effectively and demonstrate the generalizability and reliability of our findings across different types of neural network encodings and input variations.

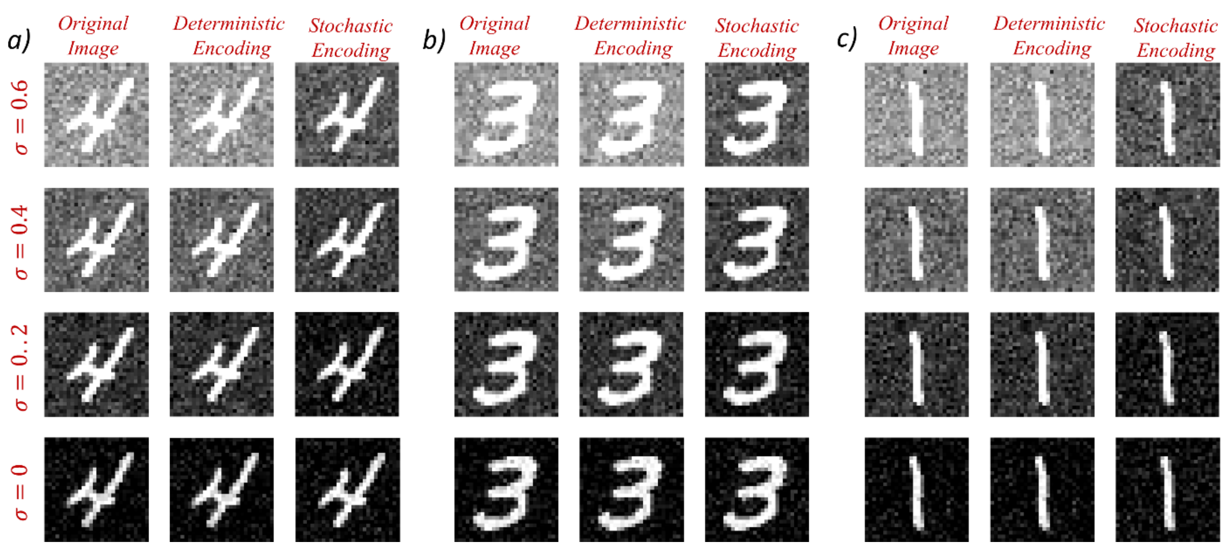

Figure R10. *Example of deterministic versus stochastic encoding of three representative MNIST handwritten digits,* ***a)*** *digit 4,* ***b)*** *digit 3, and* ***c)*** *digit 1 in the absence of noise and in presence of varying standard deviation of noise (σ = 0.2, 0.4 and 0.6). Stochastic encoding can resolve finer features in the image even in the presence of noise, compared to deterministic encoding.*

8. In Fig. 2c, the temperature condition was wrongly Annotated.

We apologize for overlooking these formatting errors. We have made the necessary corrections to Fig. 2c in the revised manuscript.

References

[1] Y.-C. Lin *et al.*, "Realizing Large-Scale, Electronic-Grade Two-Dimensional Semiconductors," *ACS Nano,* vol. 12, no. 2, pp. 965-975, 2018/02/27 2018, doi: 10.1021/acsnano.7b07059.

[2] S. Lippert *et al.*, "Influence of the substrate material on the optical properties of tungsten diselenide monolayers," *2D Materials,* vol. 4, no. 2, p. 025045, 2017/03/09 2017, doi: 10.1088/2053-1583/aa5b21.

[3] Z. Wu *et al.*, "Defect Activated Photoluminescence in WSe2 Monolayer," *The Journal of Physical Chemistry C,* vol. 121, no. 22, pp. 12294-12299, 2017/06/08 2017, doi: 10.1021/acs.jpcc.7b03585.

[4] B. Stampfer *et al.*, "Characterization of Single Defects in Ultrascaled MoS2 Field-Effect Transistors," *ACS Nano,* vol. 12, no. 6, pp. 5368-5375, 2018/06/26 2018, doi: 10.1021/acsnano.8b00268.

[5] A. Oberoi *et al.*, "Toward High-Performance p-Type Two-Dimensional Field Effect Transistors: Contact Engineering, Scaling, and Doping," *ACS Nano,* vol. 17, no. 20, pp. 19709-19723, 2023/10/24 2023, doi: 10.1021/acsnano.3c03060.

[6] J. Michl *et al.*, "Efficient Modeling of Charge Trapping at Cryogenic Temperatures—Part I: Theory," *IEEE Transactions on Electron Devices,* vol. 68, no. 12, pp. 6365-6371, 2021, doi: 10.1109/TED.2021.3116931.

[7] T. Boutchacha, G. Ghibaudo, G. Guégan, and T. Skotnicki, "Low frequency noise characterization of 0.18µm Si CMOS transistors," *Microelectronics Reliability,* vol. 37, no. 10, pp. 1599-1602, 1997/10/01/ 1997, doi: https://doi.org/10.1016/S0026-2714(97)00119-4.

[8] H. Ravichandran *et al.*, "Observation of Rich Defect Dynamics in Monolayer MoS2," *ACS Nano,* vol. 17, no. 15, pp. 14449-14460, 2023/08/08 2023, doi: 10.1021/acsnano.2c12900.

[9] D. K. Polyushkin *et al.*, "Analogue two-dimensional semiconductor electronics," *Nature Electronics,* vol. 3, no. 8, pp. 486-491, 2020/08/01 2020, doi: 10.1038/s41928-020-0460-6.

[10] S. Wachter, D. K. Polyushkin, O. Bethge, and T. Mueller, "A microprocessor based on a two-dimensional semiconductor," *Nature Communications,* vol. 8, no. 1, p. 14948, 2017/04/11 2017, doi: 10.1038/ncomms14948.

Reviewer #1 (Remarks to the Author):

Dear Authors,

Thank you for providing the additional information and making the necessary changes in the manuscript. With these updates, the goal of the study has become bit more clear. However, we still have a major comment related to the SNN training and inference accuracy of the presented SNN encoder.

We would like to thank the reviewer for their appreciation of the additional information that has been added to the manuscript and we are happy to provide further information as necessary.

One of the primary objectives of the manuscript is to demonstrate enhanced inference accuracy for noise-inflicted medical-MNIST images through a stochastic encoder. The accuracy of the inference not only depends on the presence of noise in the medical images but also on the variability of the encoder. Since the encoder in this study is characterized by the stochastic generation of spikes through WSe2 FETs, the effect of cycle-to-cycle (C2C) and device-to-device (D2D) variation on the inference accuracy is crucial.

In supplementary figure 17 you have presented the effect of Gaussian noise on inference accuracy by showing inference accuracy versus σ. However, it is unclear whether the Gaussian noise was introduced in the MNIST dataset images or in the encoding parameter τ¯_spike. The statistical study needs to be conducted with noise in the encoding parameter τ¯_spike.

If the study was indeed carried out by incorporating noise in τ¯_spike through D2D and C2C variation, then conducting the study over only three cycles and two devices is statistically insignificant. For a statistically significant study, σ should be calculated over at least 100 cycles and 10 devices.

We look forward to your response addressing this concern.

Best regards,

We appreciate the reviewer for highlighting this important aspect. We agree that understanding the impact of cycle-to-cycle (C2C) and device-to-device (D2D) variation on inference accuracy is critical. In response to the reviewer's suggestion, we conducted a statistical analysis to evaluate how D2D and C2C variations affect inference accuracy. This analysis involved performing 100 measurement cycles on a single device to obtain the C2C variations and testing 10 different devices to obtain the D2D. variations. **Fig. R1a** and **Fig. R1b** illustrate the variations in mean inter-spike interval ($\bar{\tau}_{spike}$) for different drain currents (I_{DS}) across 100 cycles on a single device and across 10 different devices, respectively. Additionally, the boxplots highlight the minimum, maximum, and mean values of $\bar{\tau}_{spike}$ for each input value of I_{DS}. To compare the distribution spread of $\bar{\tau}_{spike}$ for each input I_{DS}, we calculated the coefficient of variation (CV). For D2D variations, a maximum CV of 0.65 was observed at I_{DS} = 150 nA, and the minimum CV of 0.11 was observed at I_{DS} = 80

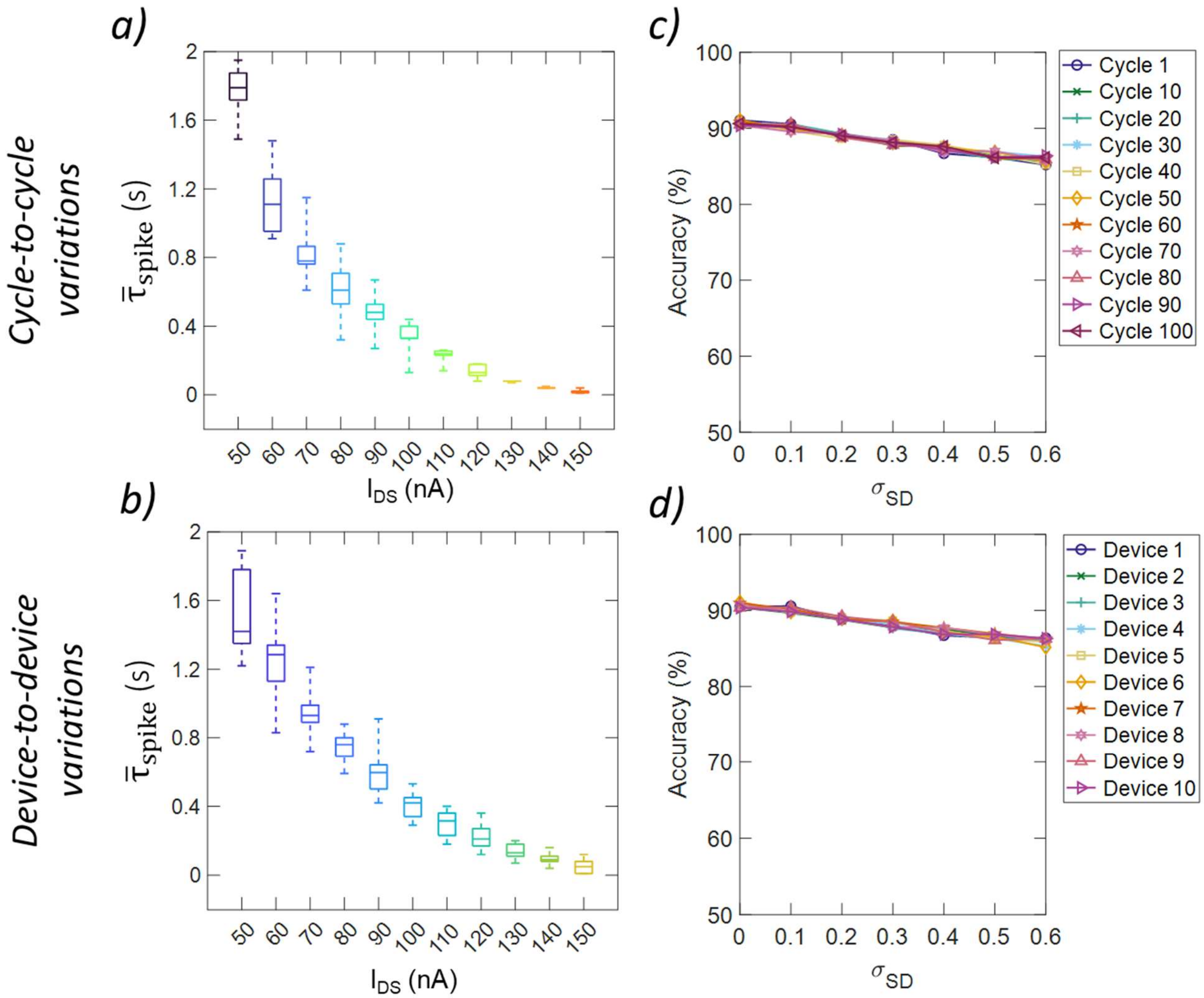

Figure R1. Impact of cycle-to-cycle and device-to-device variation on inference accuracy of medical MNIST images. *Boxplots depicting $\bar{\tau}_{spike}$ as a function of the input I_{DS} for* ***a)*** *one device over 100 cycles and* ***b)*** *ten devices for one cycle. We have used a fitting based on the empirical model given by the equation $\bar{\tau}_{spike} = \tau_0 \times exp\left(-\frac{I_0}{I_{DS}}\right)$ to quantify the relationship between $\bar{\tau}_{spike}$ and I_{DS}. Corresponding inference accuracy as a function of standard deviation (σ_{SD}) of gaussian noise added to medical MNIST images followed by stochastic encoding for* ***c)*** *cycle-to-cycle variations and* ***d)*** *device-to-device variations. The accuracy remains agnostic to cycle-to-cycle and device-to-device variations indicating the robustness of the stochastic SNN for noise-tolerant inference applications.*

nA. For C2C variations, a maximum CV of 0.45 was at I_{DS} = 150 nA, while the minimum CV of 0.08 was at I_{DS} = 50 nA. These findings indicate that $\bar{\tau}_{spike}$ is relatively consistent across cycles and devices, but inherent variations in $\bar{\tau}_{spike}$ can still impact inference accuracy. To further investigate this, we used the empirical model described in the manuscript to quantify the relationship between $\bar{\tau}_{spike}$, as given by the equation:

$$\bar{\tau}_{spike} = \tau_0 \times exp\left(-\frac{I_0}{I_{DS}}\right)$$

where I_0 and τ_0 are fitting parameters. Additionally, we introduced different levels of Gaussian noise, with standard deviations ranging from σ = 0 to σ = 0.6, to the input images. We then generated spike trains using the empirical model to represent images from the Medical-MNIST (M-MNIST) dataset. This approach allowed us to evaluate the effect of inherent noise due to C2C and D2D variations on inference accuracy. The generated spike trains, inflicted with noise, were fed into a fully trained Spiking Neural Network (SNN) to assess classification accuracy under both C2C and D2D variations, as shown in **Fig. R1c** and **Fig. R1d**. These results have been included in **Supplementary Figure 17** of the revised manuscript. Importantly, our findings suggest that the performance of our model remains unaffected despite the inherent noise in $\bar{\tau}_{spike}$ due to C2C and D2D variations and the addition of Gaussian noise to the input images. This stability could be attributed to the inherent noise tolerance of SNNs, highlighting their potential for real-world applications where variability is inevitable.

Reviewer #2 (Remarks to the Author):

The authors have addressed the questions well. I am good with publishing this version.

We would like to thank the reviewer for their appreciation of our work and recommending publication in Nature Communications.

Reviewer #3 (Remarks to the Author):

I have no more questions for the revision. I agree to accept this article for publishing on Nature Comm.

We would like to thank the reviewer for their appreciation of our work and recommending publication in Nature Communications.