Sandia National Laboratories

Exceptional service in the national interest

BIOLOGICAL DYNAMICS ENABLING TRAINING OF BINARY RECURRENT NETWORKS

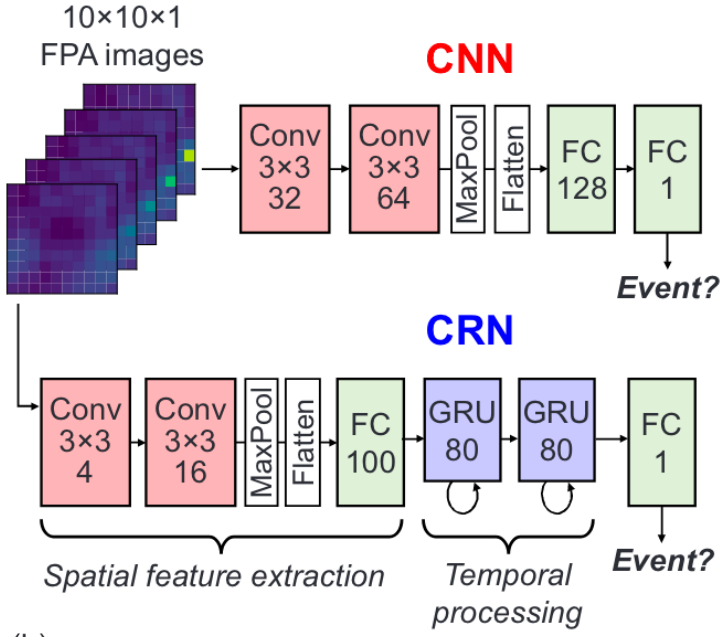
NICE 2024

G. William Chapman, Corinne Teeter, Sapan Agarwal, Patrick Xiao, Park Hays, & Srideep Musuvathy

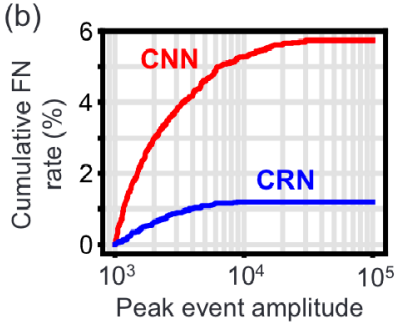
2024.04.23

SAND2024-04982C

MOTIVATION: ANALOG REMOTE SENSING



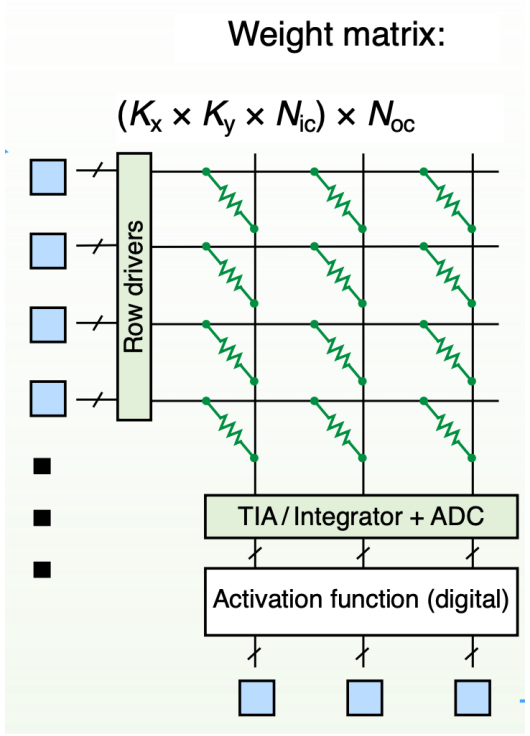
(b)



(c)

	CNN	CRN
# weights	93.2K	98.0K
Detection accuracy	96.9%	99.3%
FN rate	5.74%	1.20%
FP rate	0.54%	0.28%

- Spatial-then-temporal processing enables detection of simple waveforms

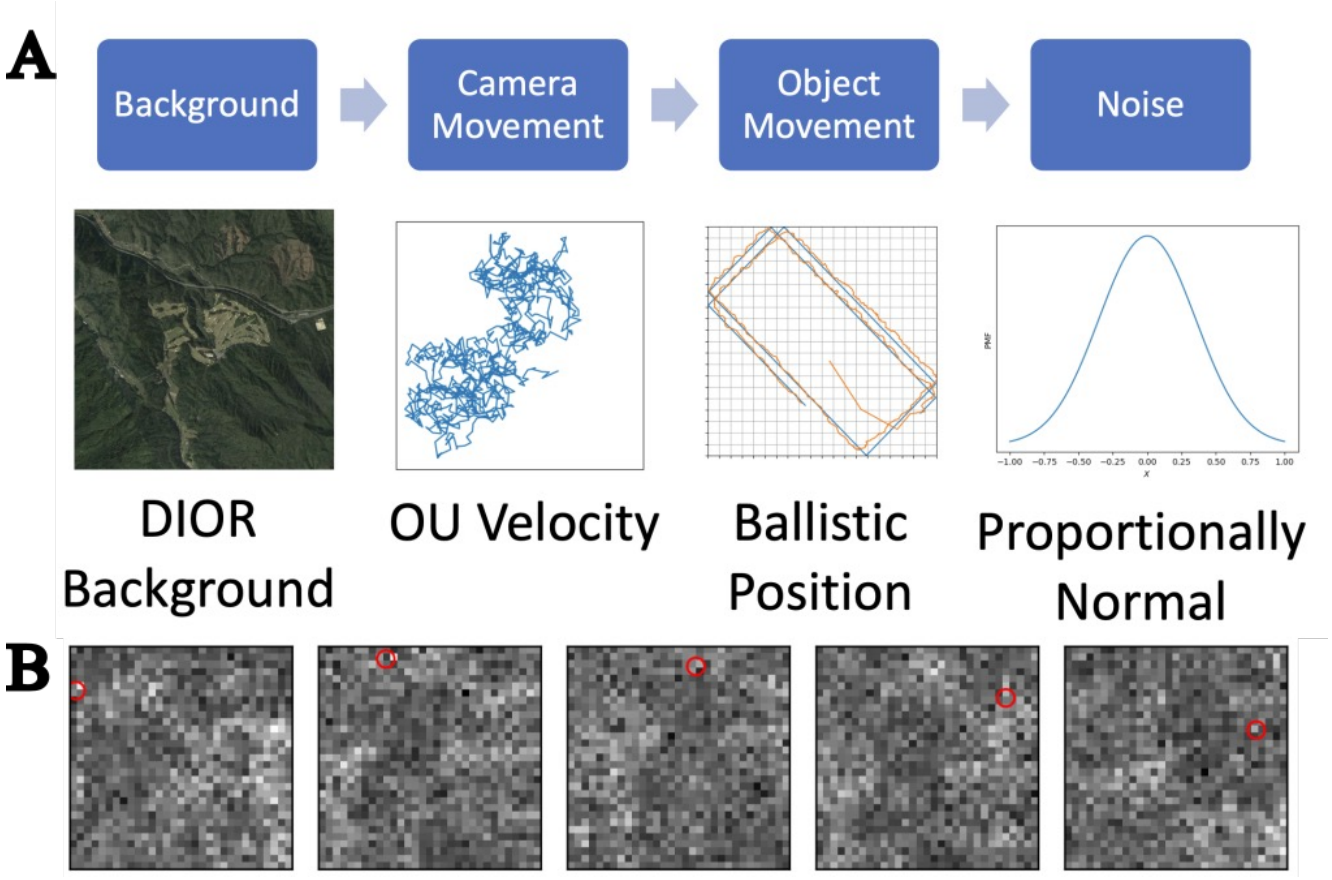


	ReLU CNN	Binary CNN
Activations	8 bits	1 bit
# Layers	4	4
# weights	20.4K	52.4k
Detection accuracy	98.80%	98.40%
Log-amplitude RMS error	0.34	0.43
Position RMS error	1.41	1.39
Processing Power	379W	26W
Min pixel pitch (processing limited)	6.2 um	1.5um
TOPS/watt	81	4180

- Vast majority of power is used by analog-to-digital conversion.
- *Binary* activations yield 50x power efficacy

Xiao et al, VLSI 2023

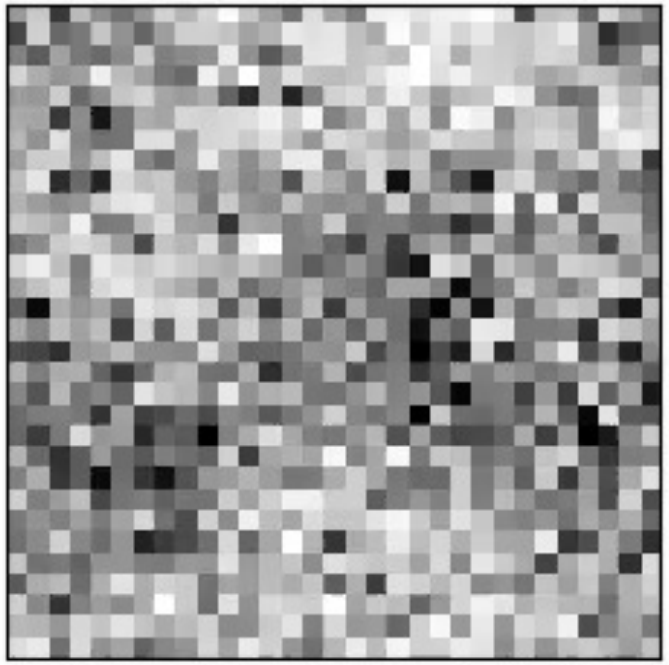
TASK GENERATION



- Adding movement, objects, and pixel-level noise allows simulation of video from still frames

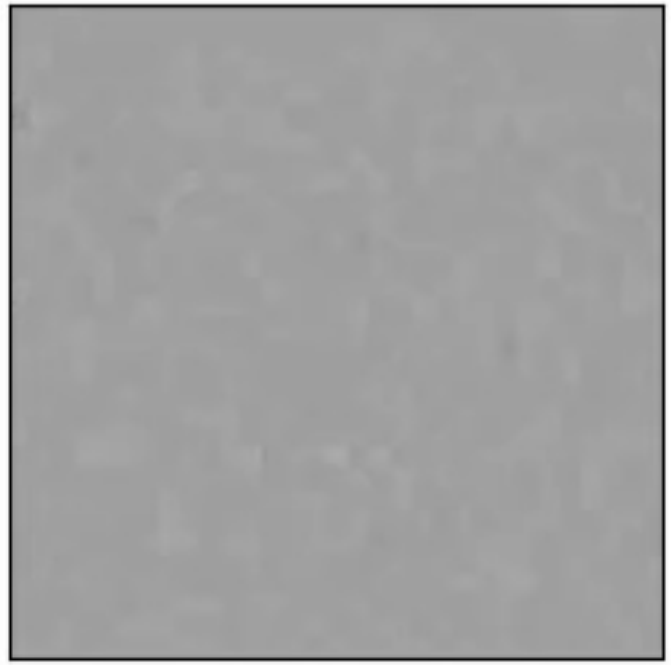
SPATIOTEMPORAL FEATURES

Input Frames



- Sub-noise intensity objects can not be localized from still images alone, but may be apparent in a video.

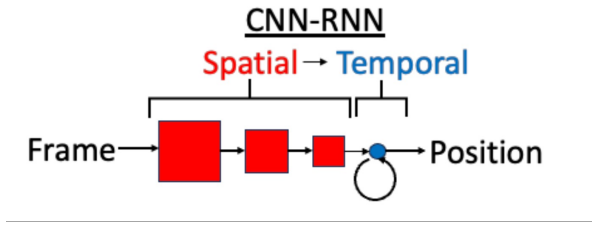
Temporal Differencing

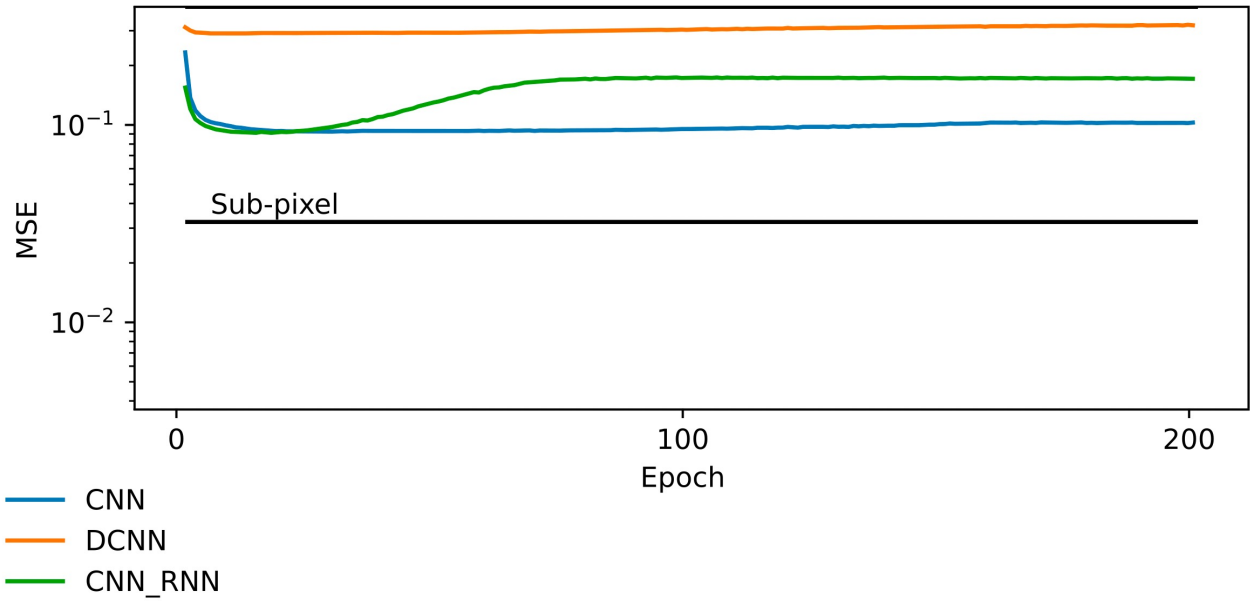


- Eg: For certain noise patterns, temporal difference alone allows object localization.
- In more complicated scenarios, additional temporal kernels are needed.

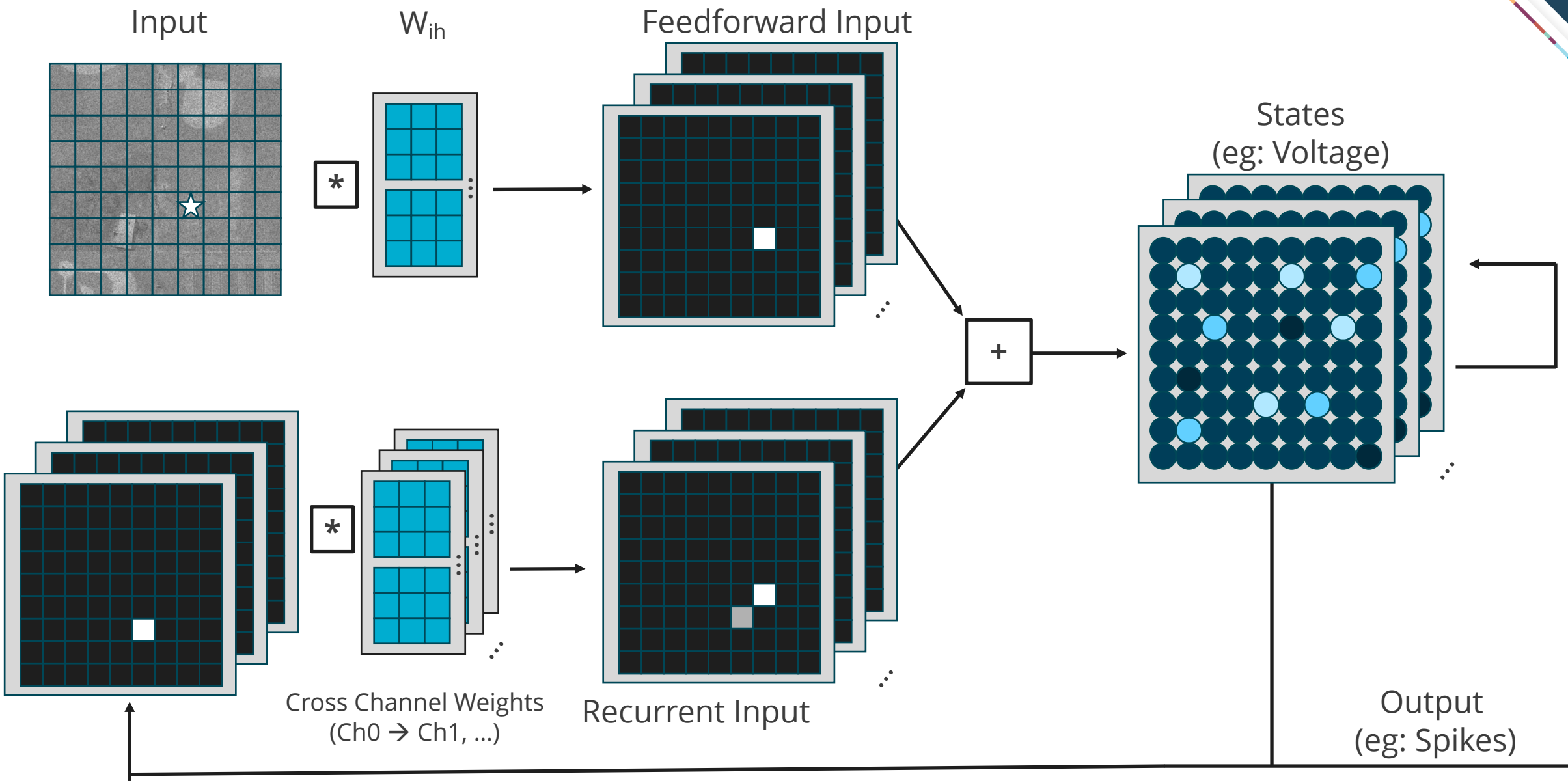
SEQUENTIAL PROCESSING PERFORMANCE

- Spatial or Spatial-then-Temporal Processing performs only better than chance.
  - Simple differencing (DCNN) is insufficient.



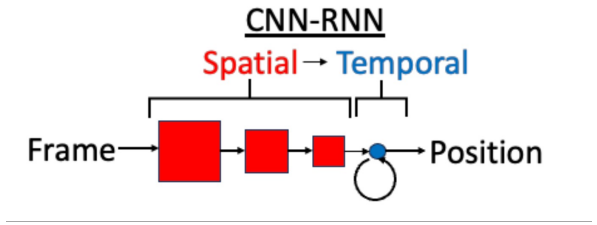


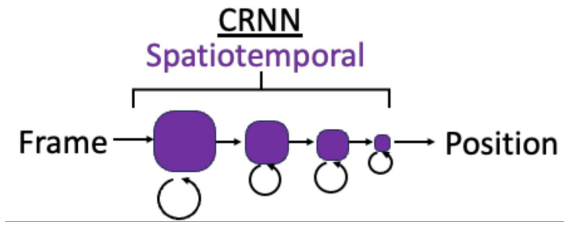
STRUCTURE OF CRNN

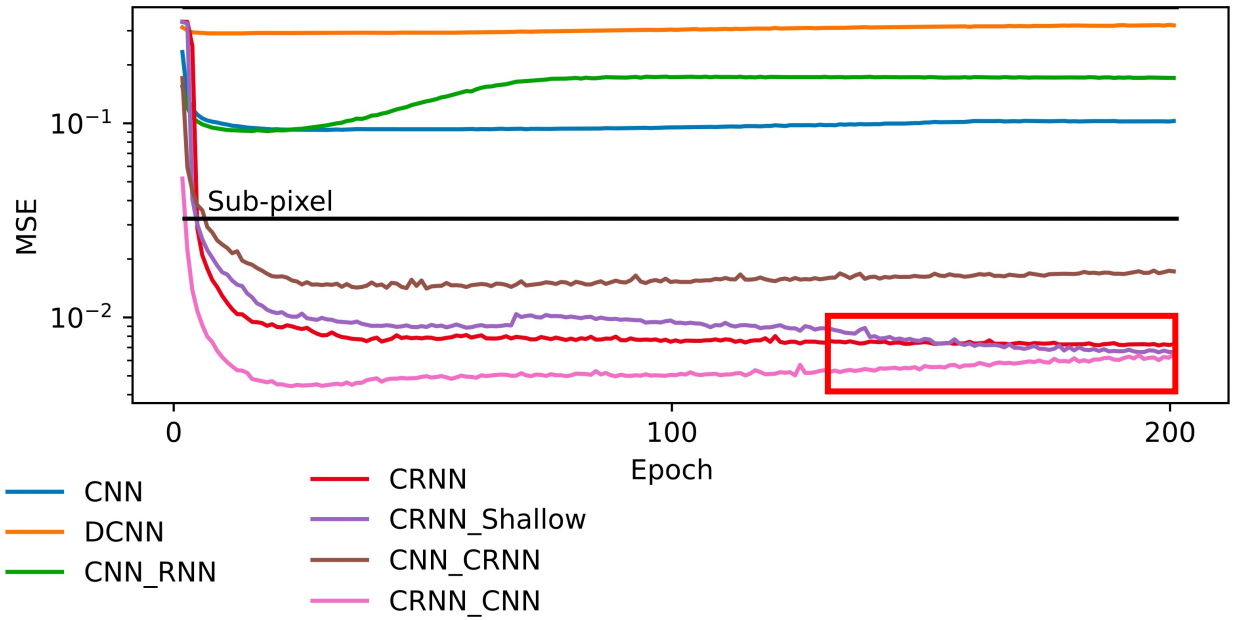


CRNN PERFORMANCE

- Spatial or Spatial-then-Temporal Processing performs only better than chance.
  - Simple differencing (DCNN) is insufficient.
- Introducing true spatiotemporal processing allows sub-pixel localization.
- A single CRNN layer *at the input* is sufficient.

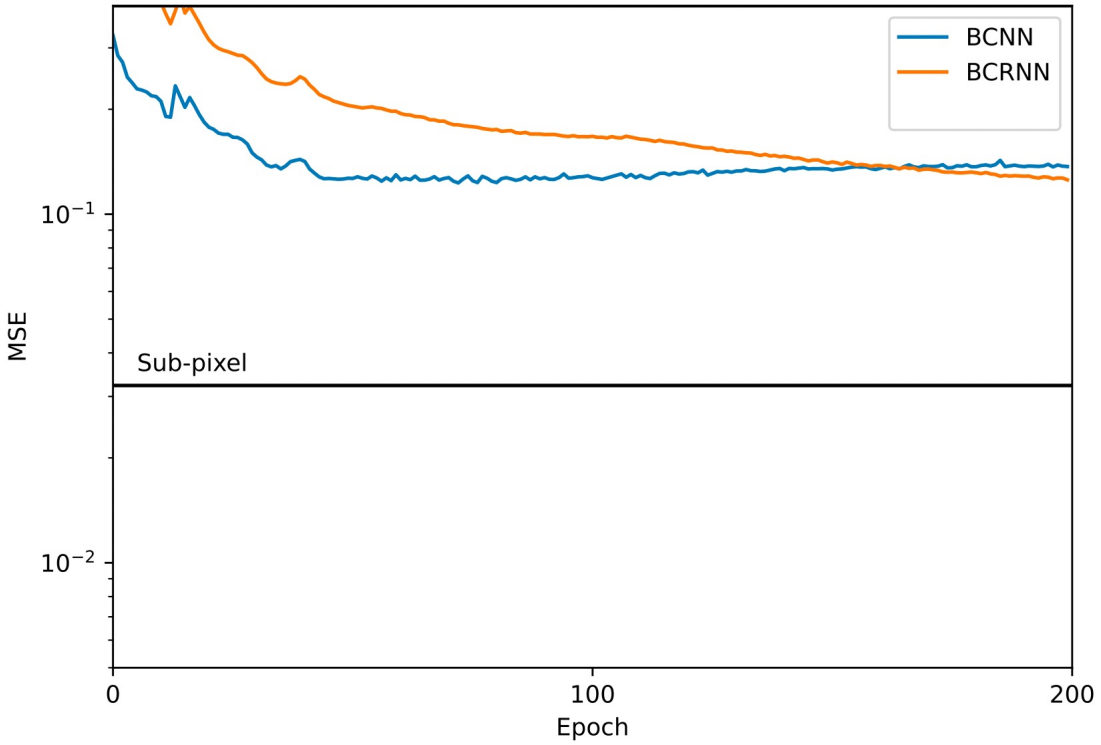


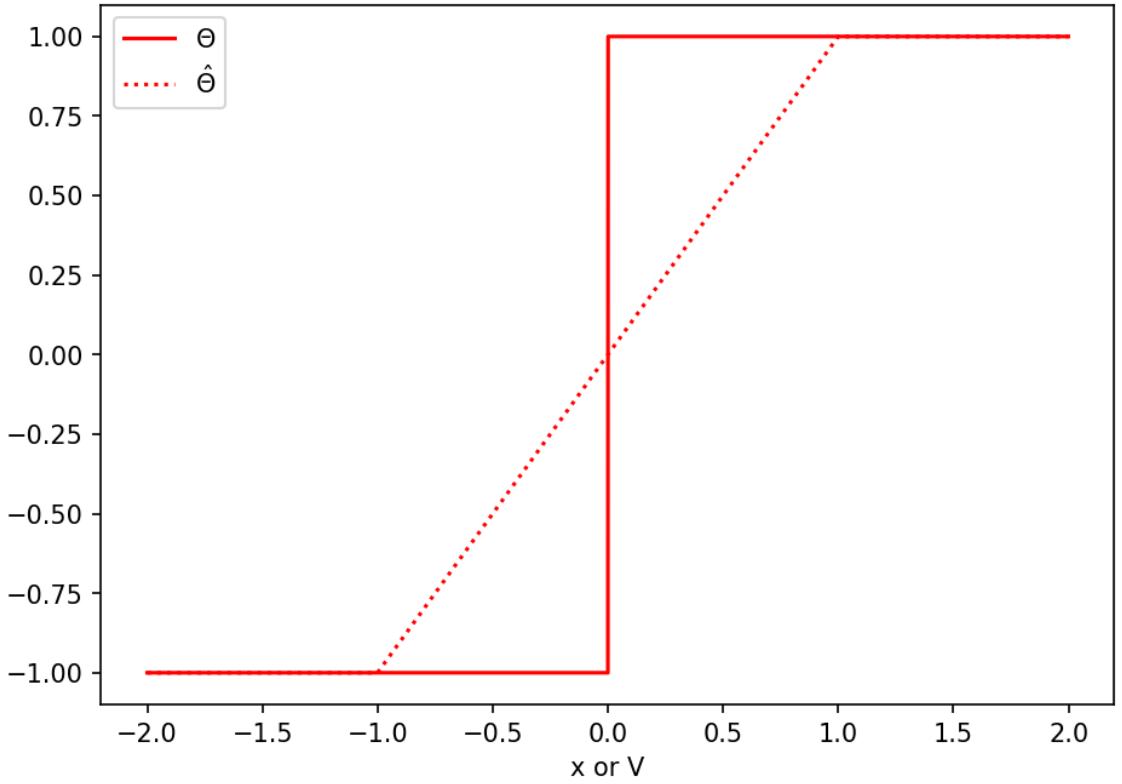




BINARY ACTIVATION RECURRENT NETWORKS

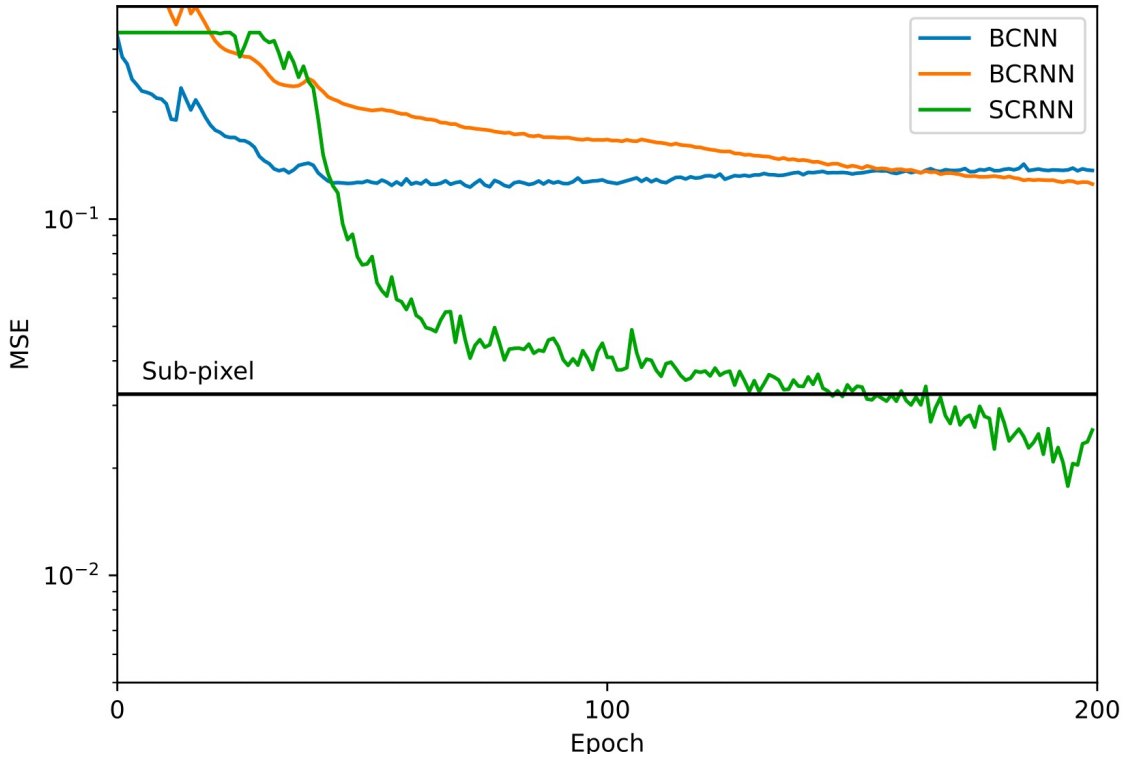
- Problem: Binary recurrent networks fail to converge over learning
- Surrogate gradient descent approximations are invalid

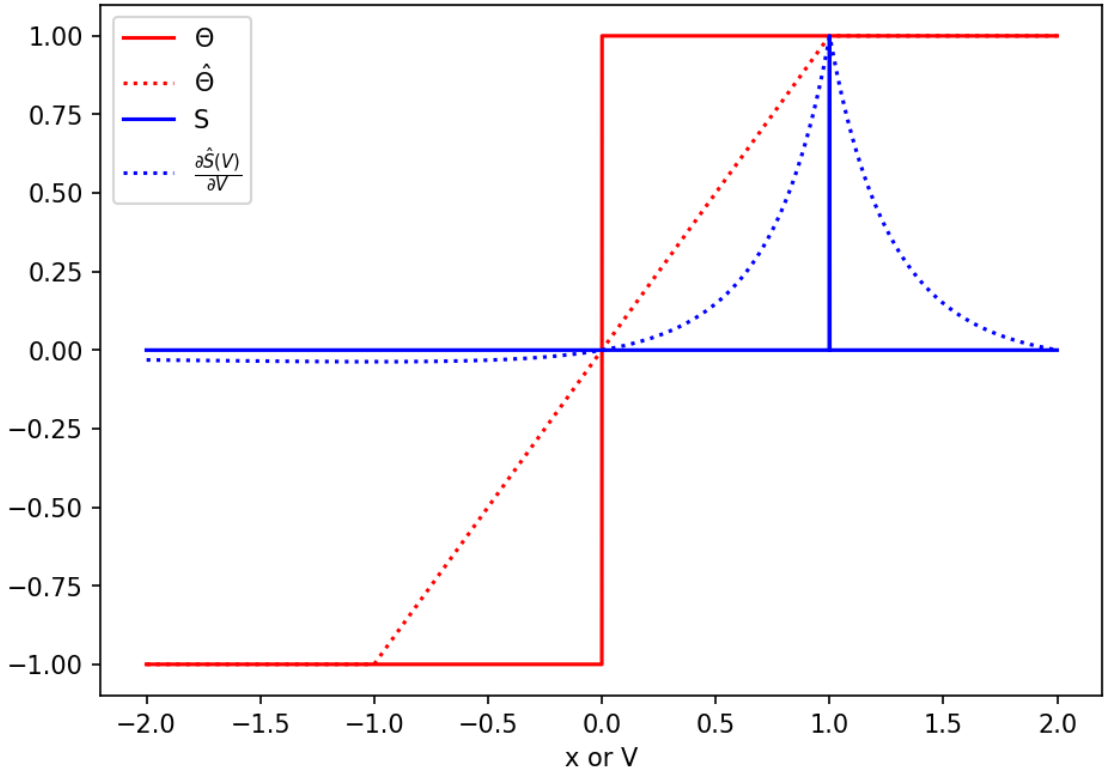




BINARY ACTIVATION RECURRENT NETWORKS

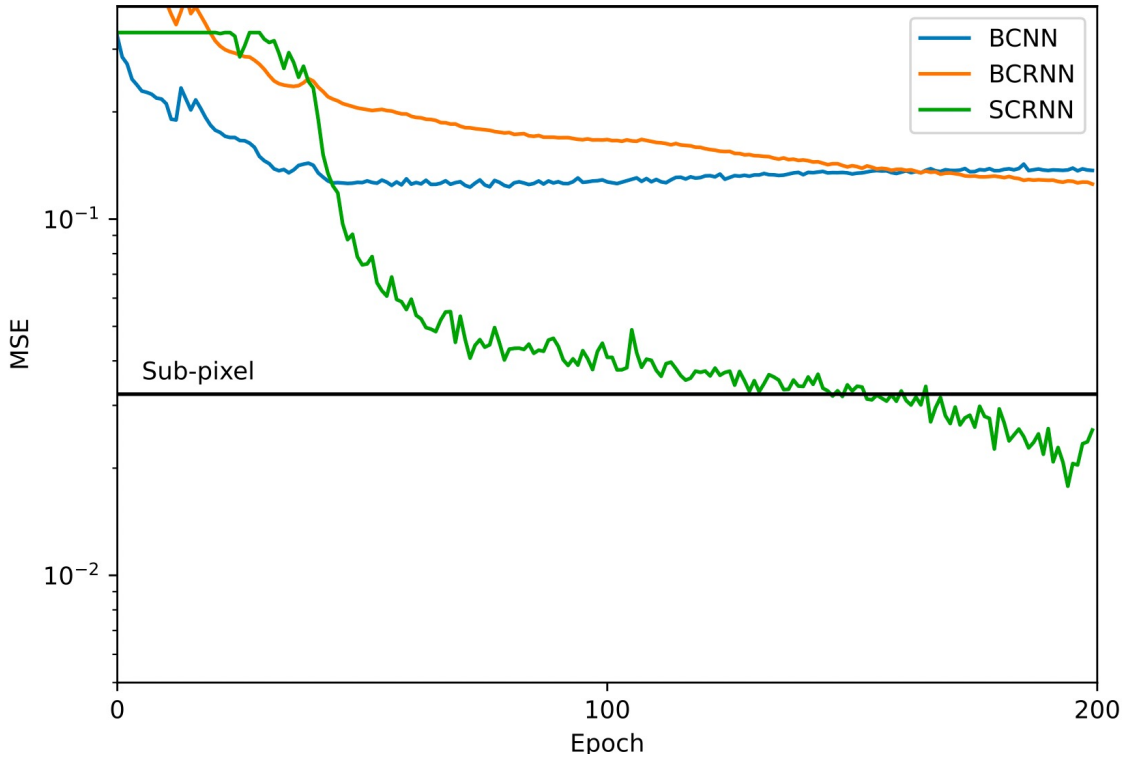
- Problem: Binary recurrent networks fail to converge over learning
- Surrogate gradient descent approximations are invalid



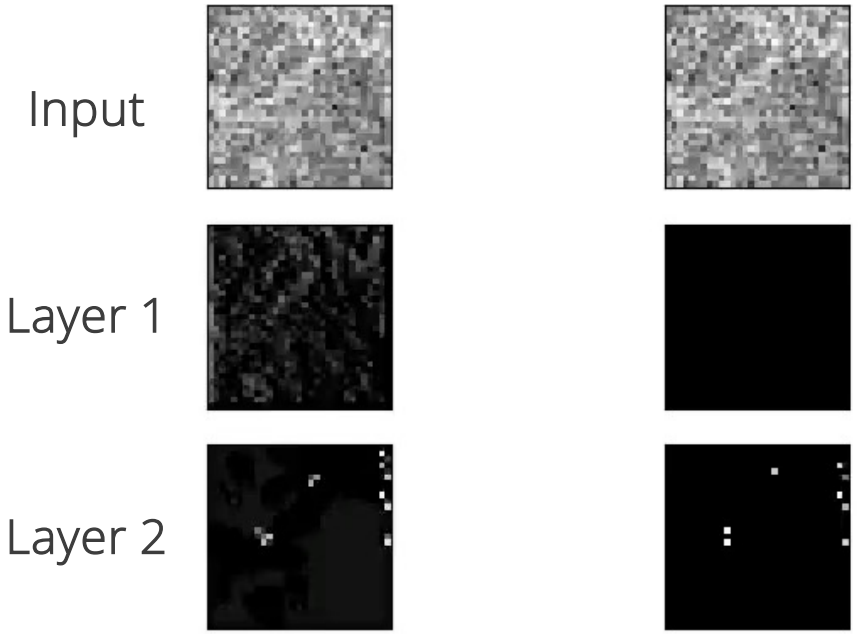


BINARY ACTIVATION RECURRENT NETWORKS

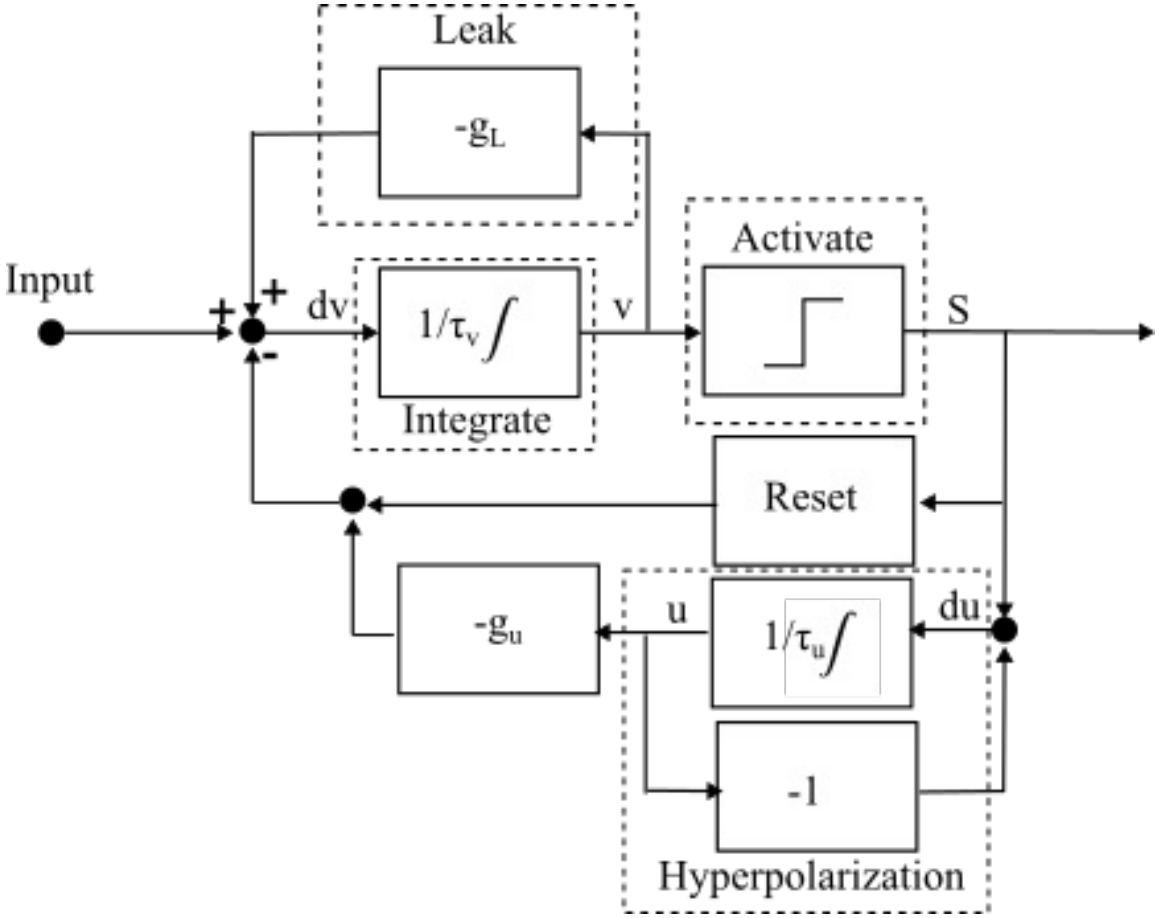
- Problem: Binary recurrent networks are unstable in time.
  - Surrogate gradient descent approximations are invalid



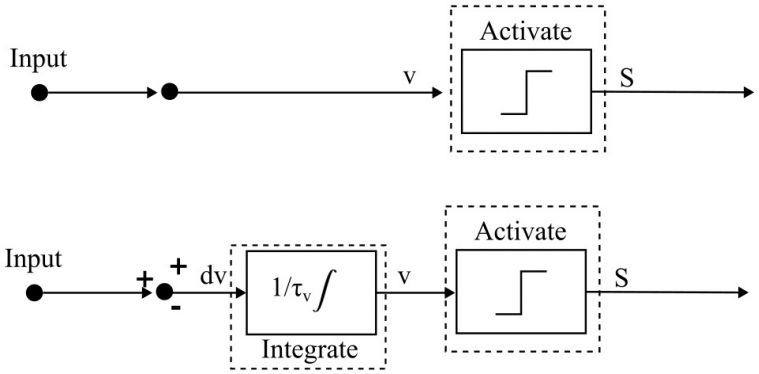
- Neuro-Solution: LIF networks can be trained with BPTT, but why?

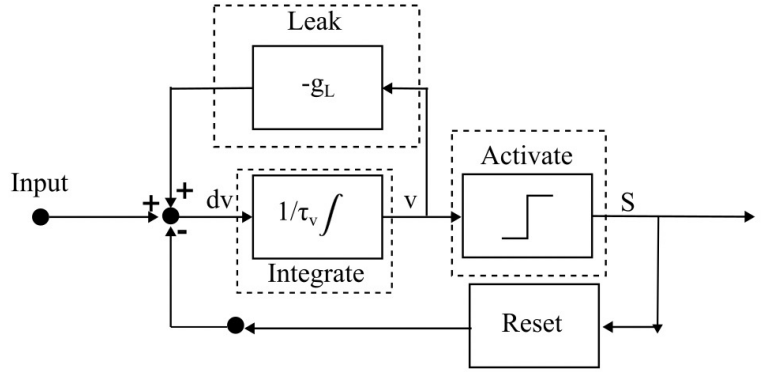

Input
Layer 1
Layer 2

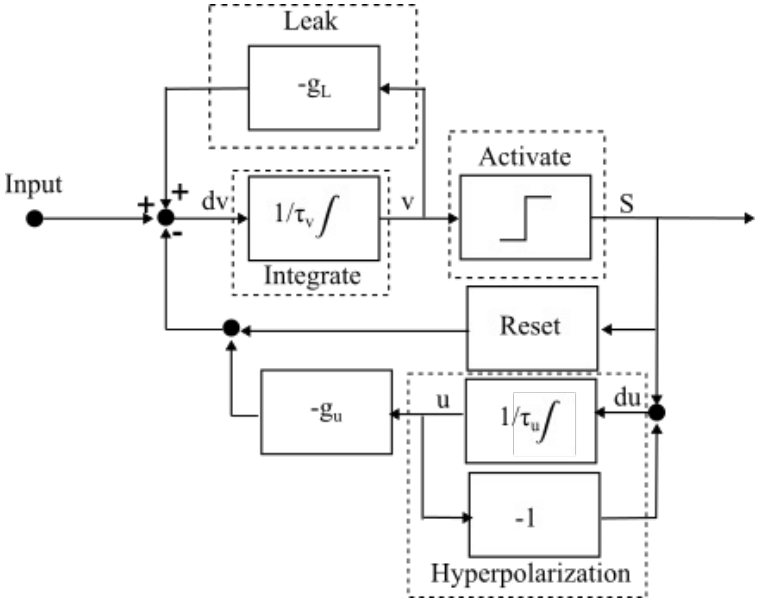
GENERALIZED LEAKY INTEGRATE AND FIRE (GLIF)



- Integrate: Temporal Low-pass filter
- Activation: Heaviside
- Leak: Decay to rest
- Reset: Explicit return to rest after threshold.
- Hyperpolarization: Post-spike inhibition

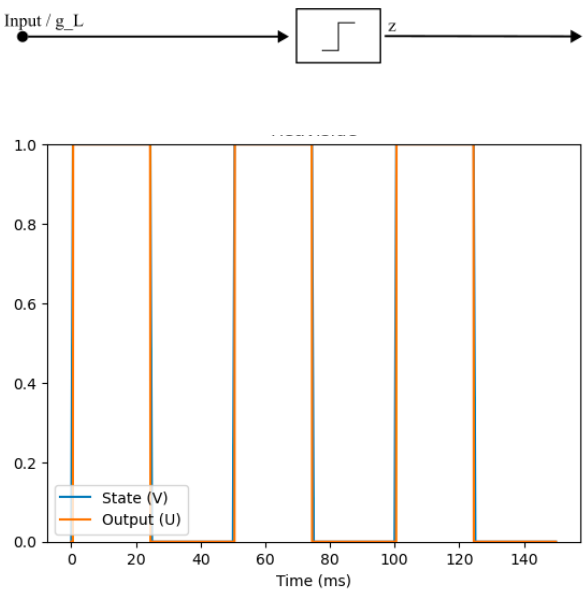

Input
v
Activate
S
Input
+
+
-
dv
1/τv ∫
Integrate
v
Activate
S


Leak
-gL
Input
+
+
-
dv
1/τv ∫
Integrate
v
Activate
S
Reset

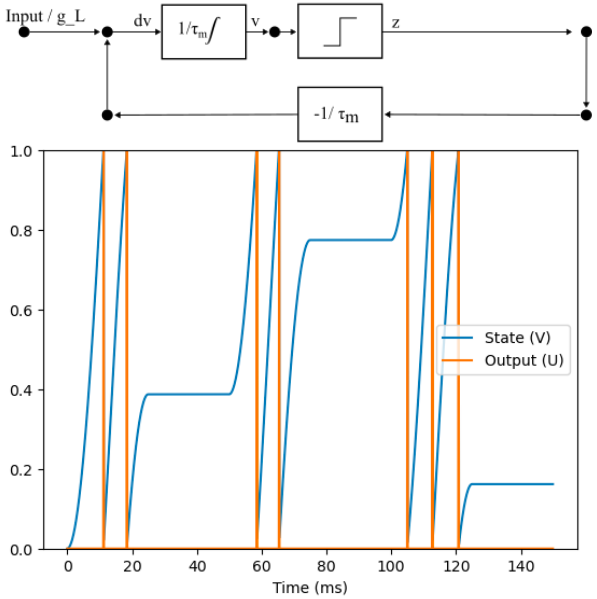

Leak
-gL
Input
+
+
-
dv
1/τv ∫
Integrate
v
Activate
S
Reset
-gu
u
1/τu ∫
du
-1
Hyperpolarization

GLIF DYNAMICS

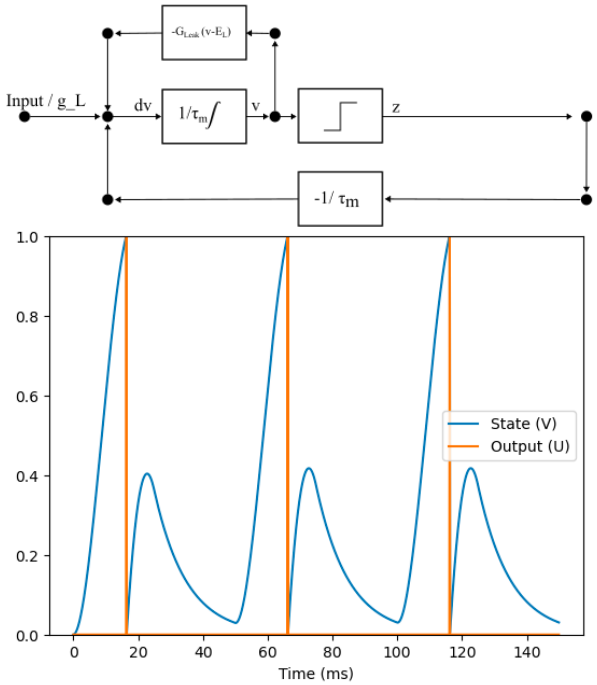
Binary Activation



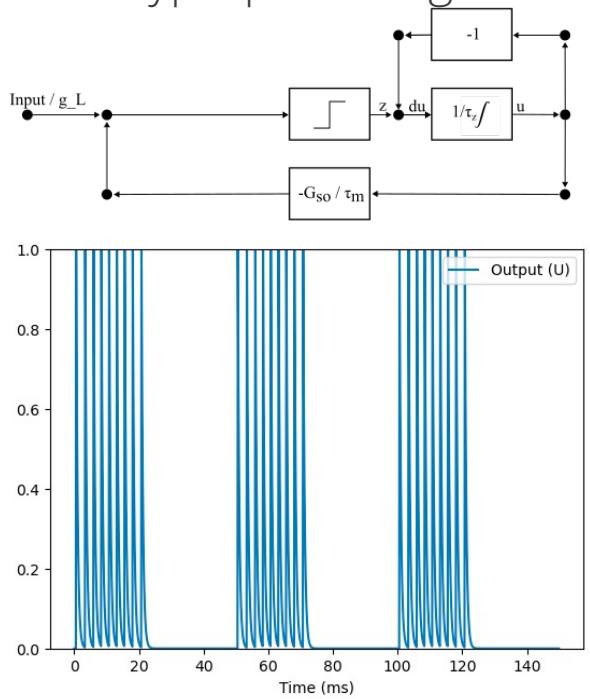
IAF



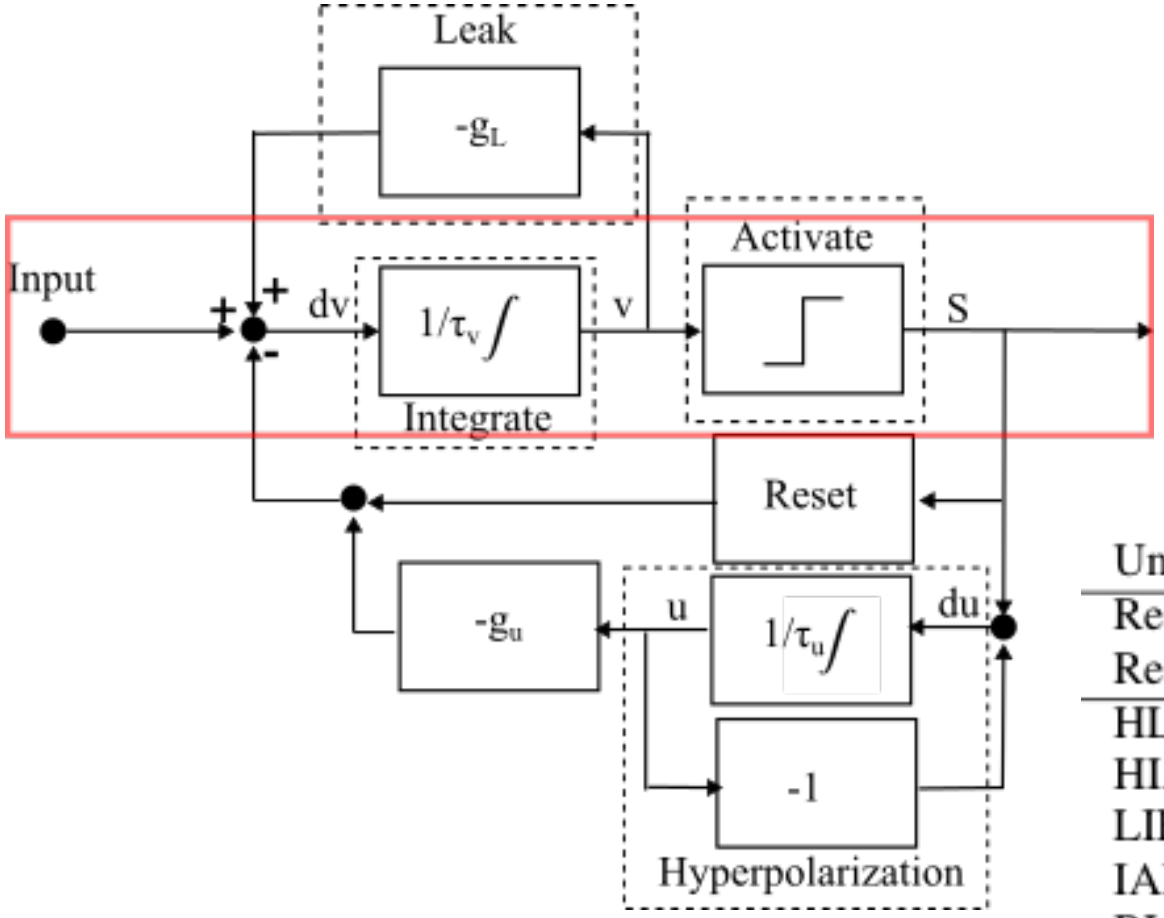
LIF



Hyperpolarizing BA



STATE IS ESSENTIAL FOR TRAINING

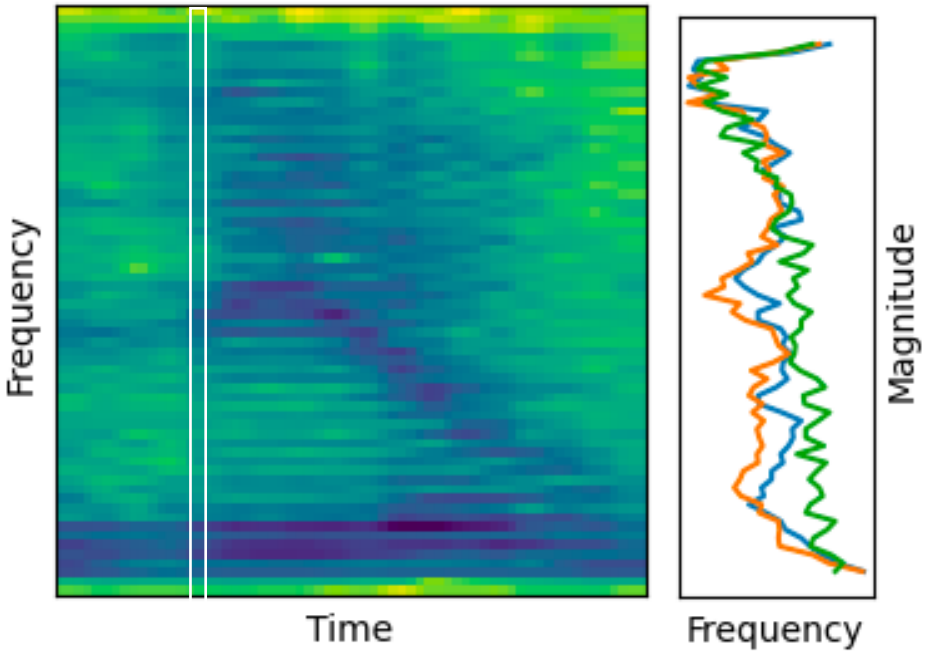


- **Integrate: Temporal Low-pass filter**
- Activation: Heaviside
- Leak: Decay to rest
- Reset: Explicit return to rest after threshold.
- Hyperpolarization: Post-spike inhibition

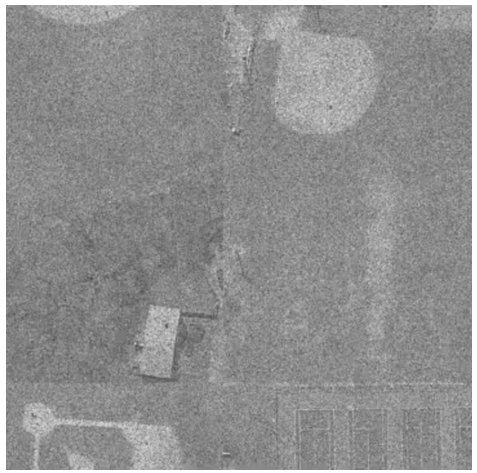
Unit Type	SOT (MSE)
Real (CNN-RNN)	.091
Real	.007
HLIF	.015
HIAF	.018
LIF	.017
IAF	.016
BLI	**.011**
BI	.013
HBA	.310
BA	.123

- Conclusion: Integrative state is responsible for enabling training.
- Guarantees smooth temporal signal for surrogate BPTT.

MULTI-TASK COMPARISON



Free-Spoken Digits Dataset (FSDD) (Temporal)



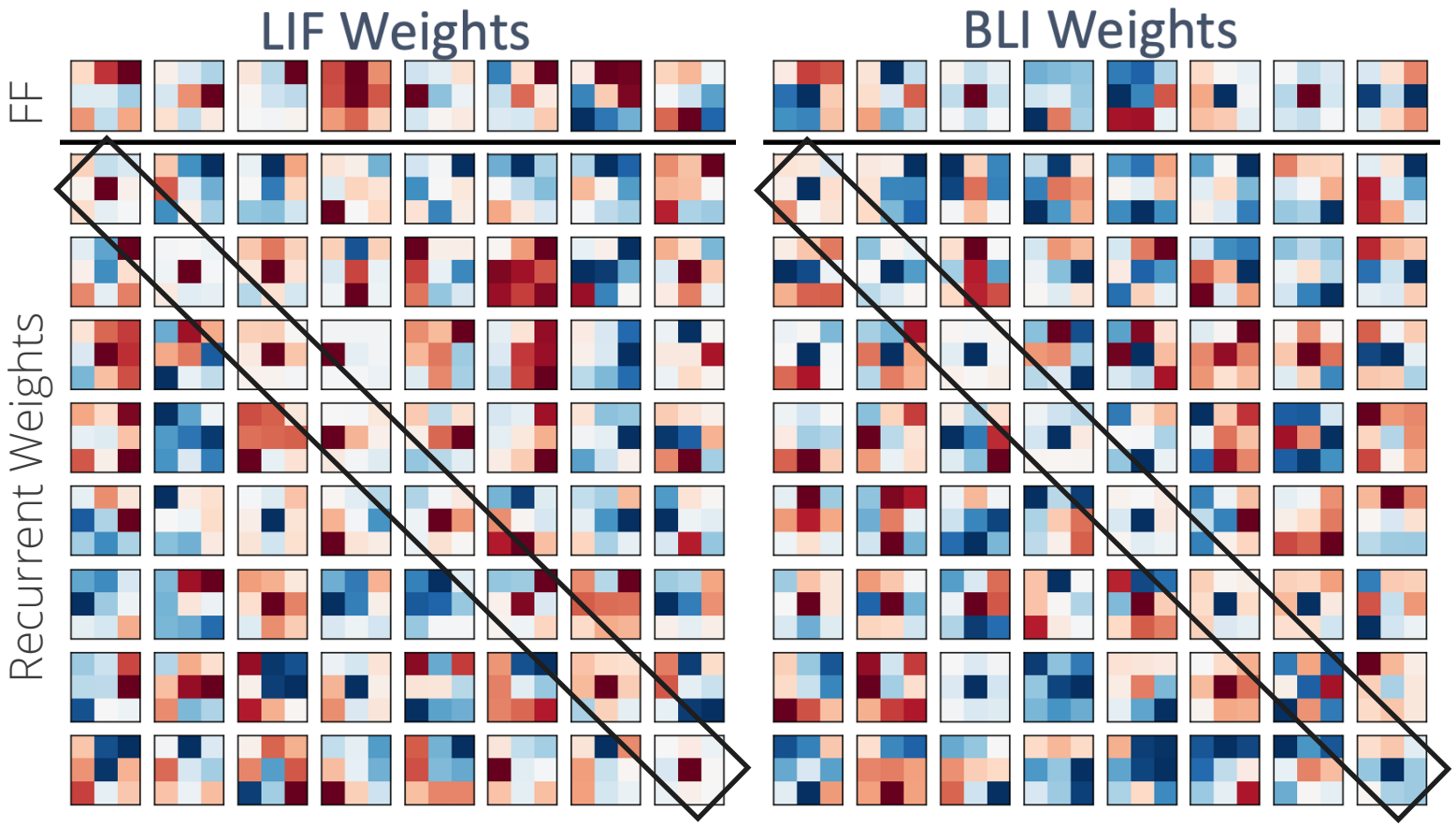
Small-Object Tracking (SOT) (Spatiotemporal)

MNIST (Spatial)

- Binary Leaky Integrator consistently performs best, across multiple task domains.

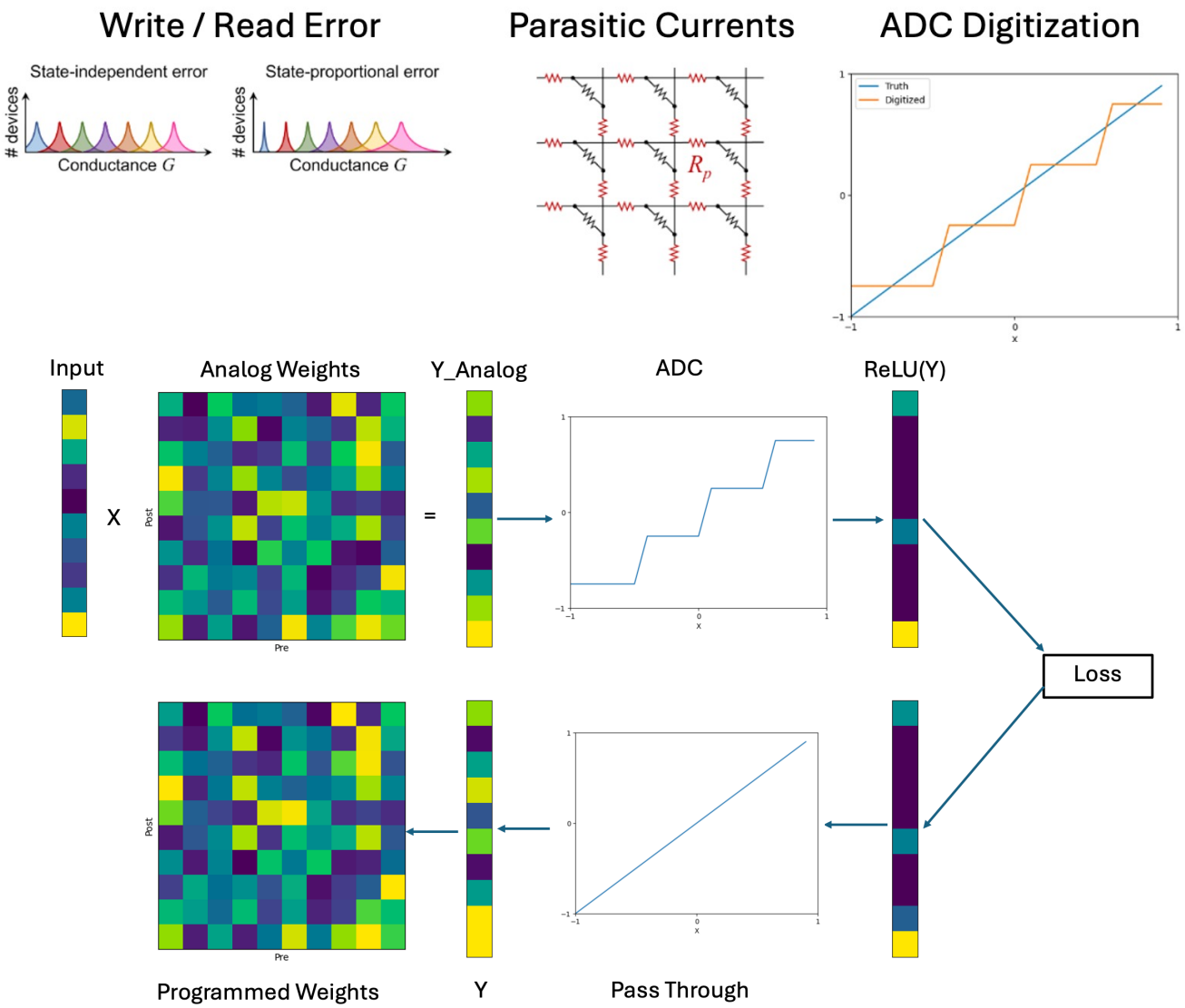
Unit Type	MNIST	FSDD	SOT (MSE)
Real (CNN-RNN)	98.9	90.4	.091
Real	99.0	98.2	.007
HLIF	98.2	95.6	.015
HIAF	98.4	94.3	.018
LIF	98.5	93.1	.017
IAF	**98.7**	92.7	.016
BLI	**98.7**	**97.2**	**.011**
BI	98.5	93.2	.013
HBA	98.5	31.8	.310
BA	97.2	48.6	.123

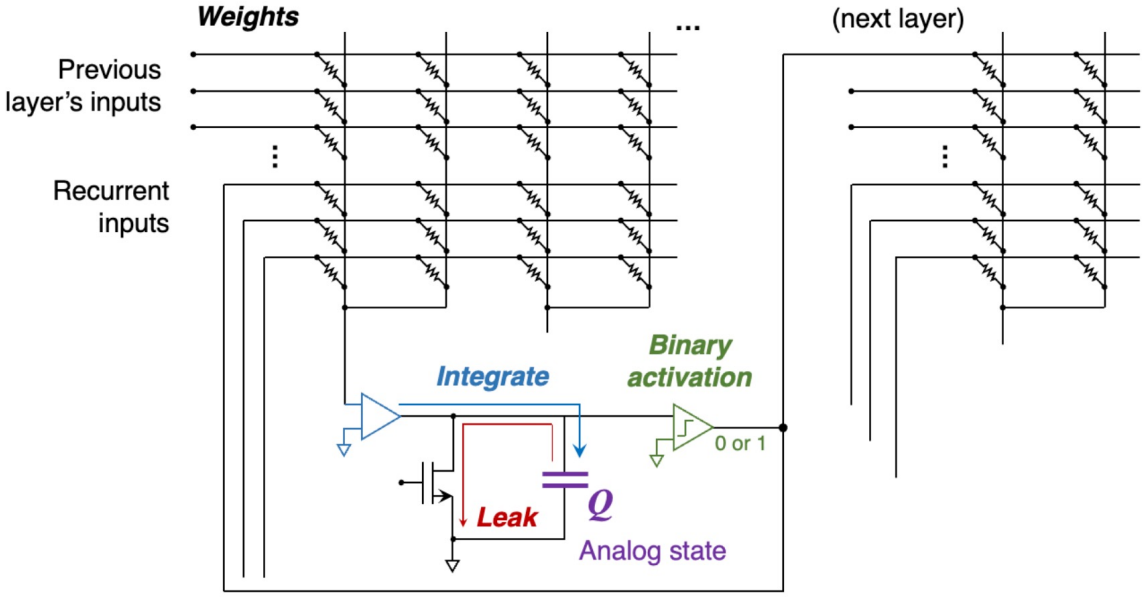
NON-SPIKING UNITS LEARN WEAK RESETTING MECHANISM



- Spiking units learn to-self excite in all three tasks.
- Non-spiking units learn to self-inhibit, preventing sustained activation.
- Post-activation potential is similar (0.2)
- Similar overall activation, with ~10% active on each frame.

ANALOG SIMULATION





- Analog-in-the-Loop Training enables *analog* spiking CRNNs to perform these tasks.
- CrossSim Tutorial: Friday, Slot 3.
  - Implementing spiking neural networks on analog hardware with training in the loop.

SUMMARY

- Convolutional-Recurrent processing allows detection of low-SNR objects by spatiotemporal pattern detection.
- LIF units support CRNN network training, with low precision activation.
- Binary-leaky Integrators contain the minimal complexity for training.
  - *Via surrogate gradient descent for ML-like tasks*.
- Analog SNNs can be trained for the same tasks, closing the co-design loop.