Kanerva++: extending the Kanerva machine with differentiable, locally block allocated latent memory.

ICLR 2021

Jason Ramapuram, Yan Wu, Alexandros Kalousis
jramapuram@apple.com, yanwu@google.com, alexandros.kalousis@hesge.ch

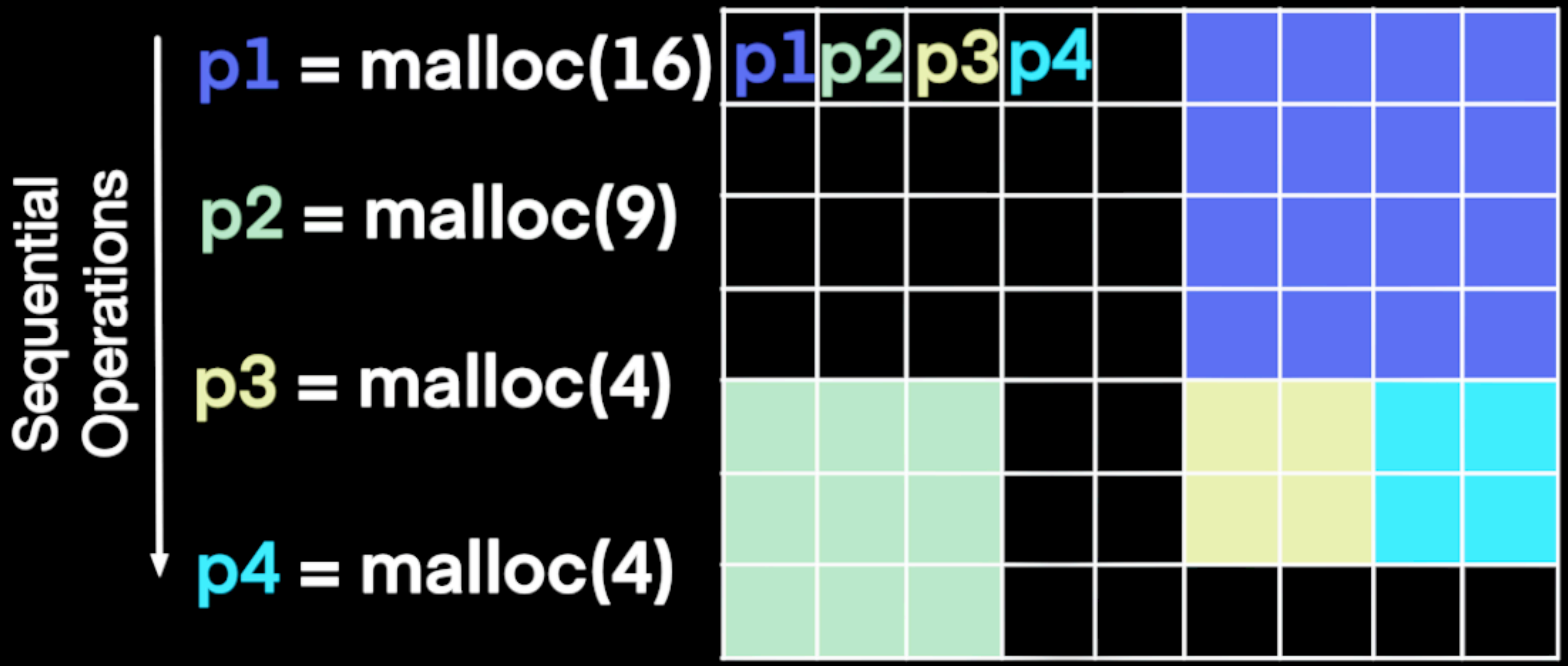
Kanerva machine & Dynamic Kanerva machine [3,4]

$$M \sim p(M) = \mathcal{N}(U, R, C)$$

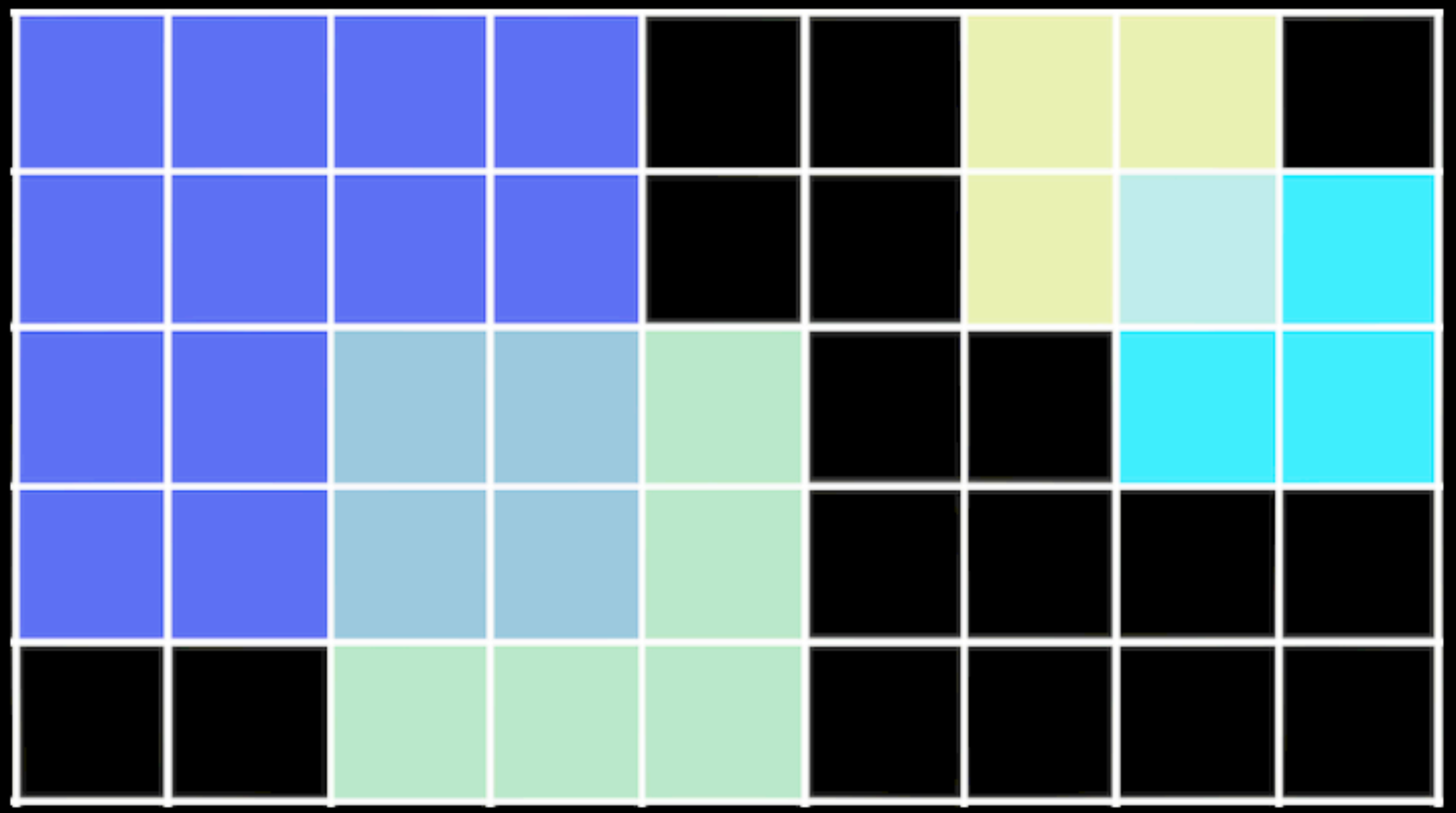
- Memory treated as a distribution.
- Writes are inference: $p(M \mid X)$
- Reads are sampling conditional: $p(X \mid M)$

[3] Wu, Yan, et al. "The Kanerva Machine: A Generative Distributed Memory." ICLR. 2018.
[4] Wu, Yan, et al. "Learning attractor dynamics for generative memory." *Advances in Neural Information Processing Systems*. 2018.

Heap allocator vs. Kanerva++

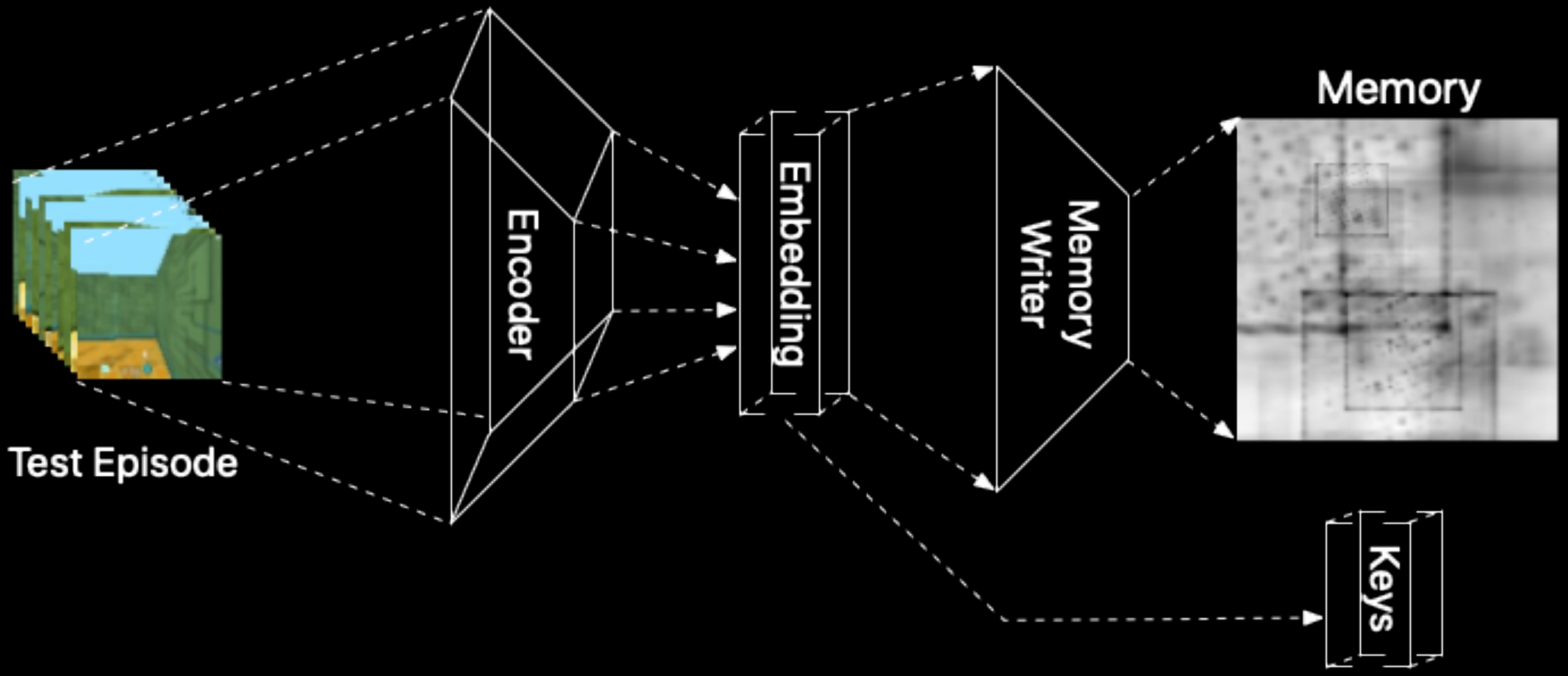


Traditional allocator



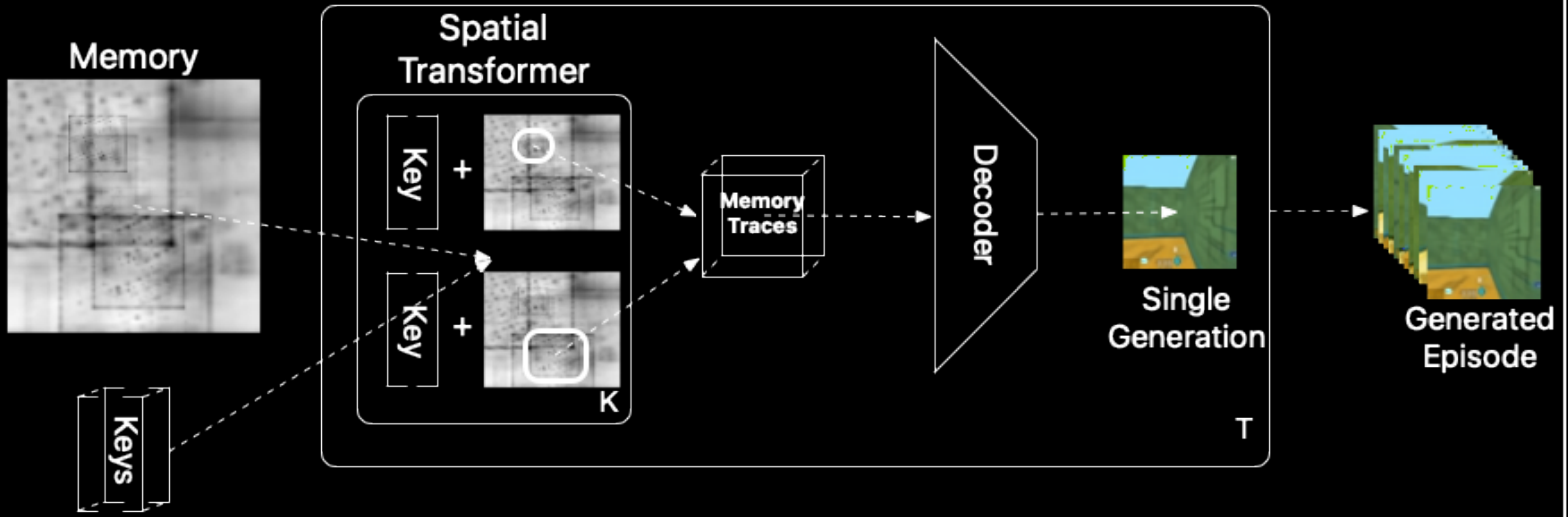
K++ allocator

Write model.



sample episode:
$X_t = \{x_1, ..., x_T\} \sim \mathcal{D}$;
compute embedding:
$E = f_{\theta_{enc}}(X_t)$;
infer keys: $Y_t \sim$
$\mathcal{N}(\mu_{\theta_{key}}(E), \sigma^2_{\theta_{key}}(E))$;
write memory:
$M \sim \delta(f_{\theta_{mem}}(E))$

Generative model.



given memory: M ;

sample keys: $Y_t \sim P(Y_t)$;

extract regions: $\hat{M} = \{f_{\theta_{ST}}(M, y_{tk})\}_{k=1}^{K}$;

infer latent: $Z_t \sim \mathcal{N}(\mu_{\theta_Z}(\hat{M}), \sigma^2_{\theta_Z}(\hat{M}))$;

decode: $\hat{X}_t \sim \mathcal{N}(\mu_{\theta_{dec}}(\mu_{Z_t}), \sigma^2)$

Optimization objective.

$$\mathcal{L}_T \approx \mathbb{E}_{q_\phi(Z|X),\, q_\phi(Y|X)} \ln p_\theta(X|Z, \hat{M}, Y)$$

$$-$$

$$\underbrace{\mathbb{E}_{q_\phi(Y|X)} \mathcal{D}_{KL}[q_\phi(Z|X) || p_\theta(Z|\hat{M}, Y)]}_{\text{Amortized latent variable posterior vs. learned prior}}$$

$$-$$

$$\underbrace{\mathcal{D}_{KL}[q_\phi(Y|X) || p(Y)]}_{\text{Amortized key posterior vs. key prior}}$$

Deterministic memory

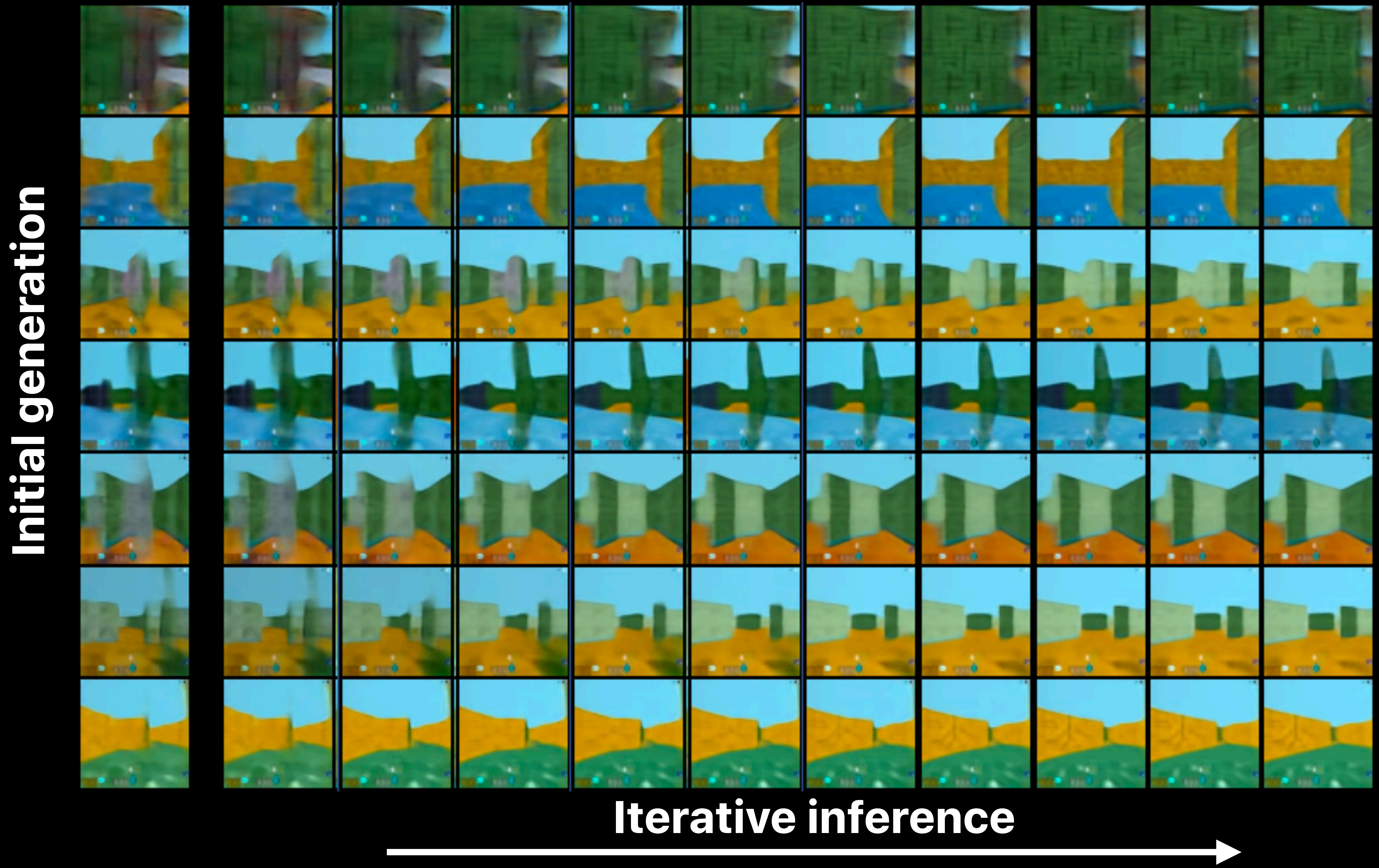
$$\hat{M} \sim \delta\left[\{f_{ST}(M, y_{tk})\}_{k=1}^{K}\right]$$

Results: memory conditional likelihood estimation

$$\ln p(X|M) \geq \mathscr{L}_T$$

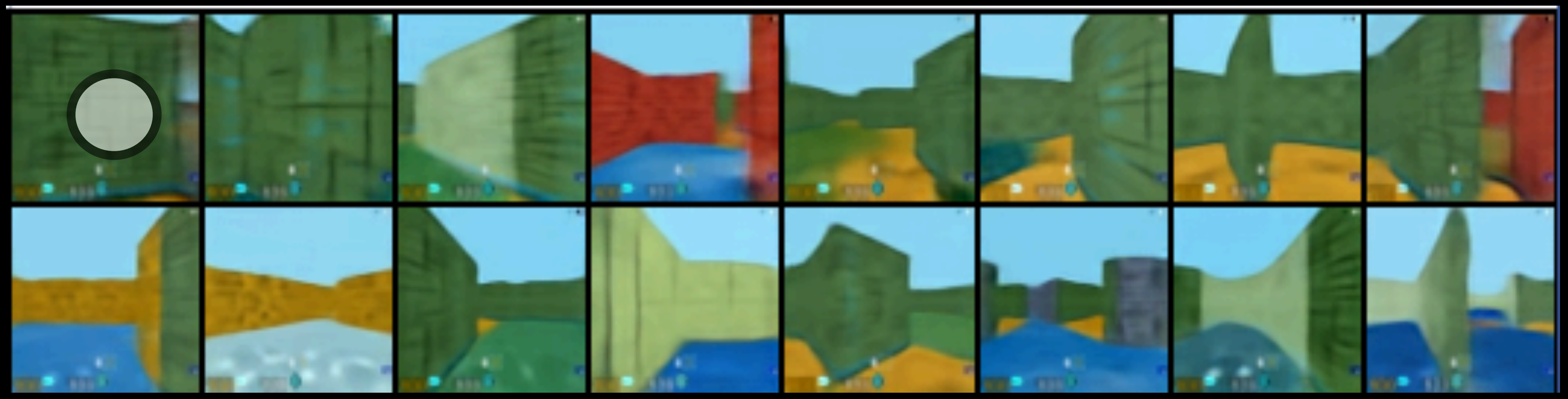
	Binarized MNIST (nats/image)	Binarized Omniglot (nats/image)	Fashion MNIST (bits/dim)	CIFAR10 (bits/dim)	DMLab mazes (bits/dim)
VMA (Bornschein, 2017)	-	103.6	-	-	-
KM (Wu, 2018a)	-	$\leq$ 68.3	-	$\leq$ 4.37	-
DNC (Graves, 2016)	-	$\leq$ 100	-	-	-
DKM (Wu 2018b)	$\leq$ 75.3	$\leq$ 77.2	-	$\leq$ 4.79	$\leq$ **2.75**
DKM w/ TSM (our impl)	$\leq$ 51.84	$\leq$ 70.88	$\leq$ 4.15	$\leq$ 4.31	$\leq$ 2.92
K++ (ours)	$\leq$ **41.58**	$\leq$ **66.24**	$\leq$ **3.4**	$\leq$ **3.28**	$\leq$ 2.88

Improving generations through iterative inference.



DMLab maze generation.

$$y_t + \epsilon$$
$$y_t \sim p(Y),\ \epsilon \sim \mathcal{N}(0, 0.01)$$



$$y_t \sim p(Y)$$

Poster #1802

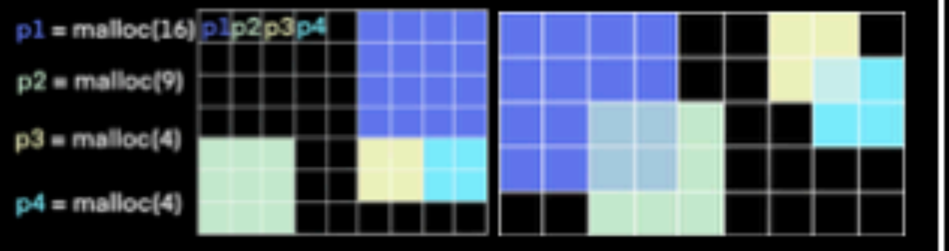
May 6, session 10

Kanerva++: extending the Kanerva Machine with differentiable, locally block allocated latent memory.

Jason Ramapuram, Yan Wu, Alexandros Kalousis
jramapuram@apple.com, yanwu@google.com, alexandros.kalousis@hesge.ch

Memory conditional generative model

- Heap allocator inspired differentiable memory.

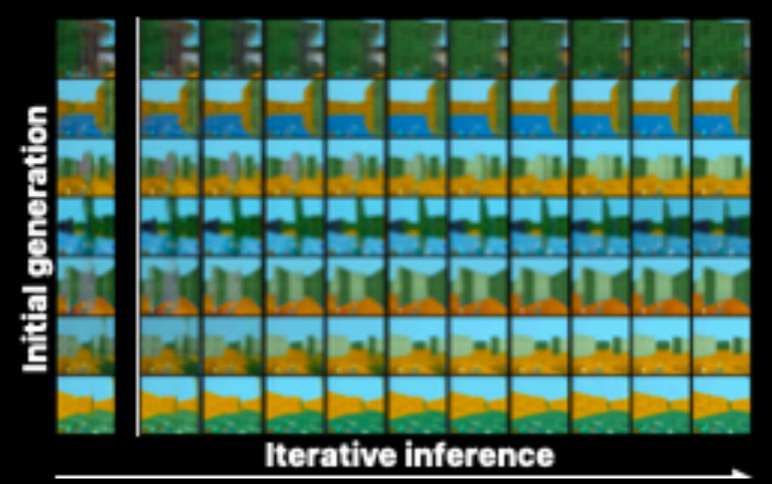


We propose

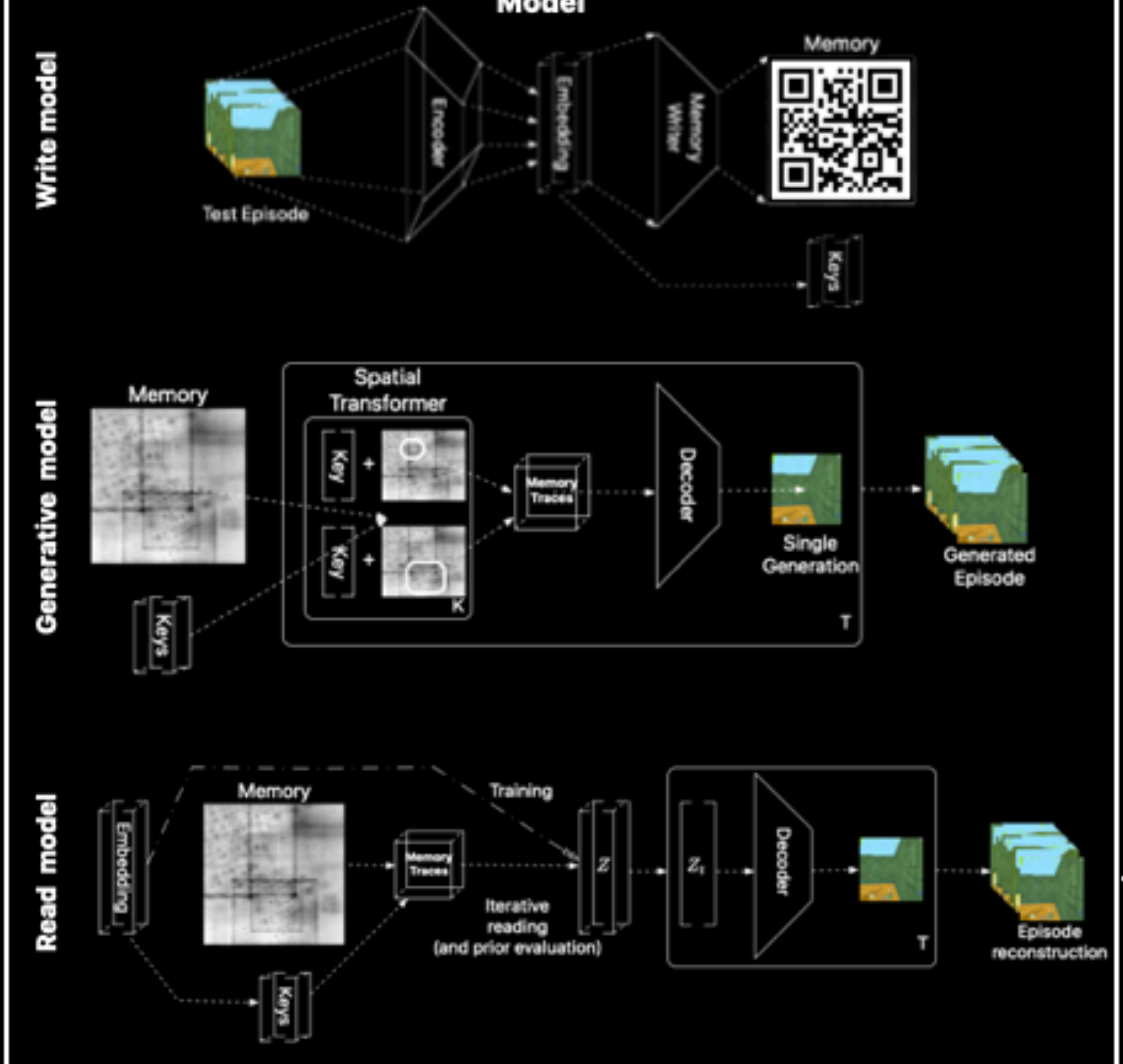
Kanerva++: a differentiable, **end-to-end** block allocated episodic memory for conditional image generation.

Contributions

- Deterministic **episodic memory** created through temporal shift module encoder.
- Spatial transfomer based **block allocation.**
- **Low dimensional sampling distribution** ($\mathbb{R}^3$) that indexes large memory.
- Local key perturbations produce **semantically meaningful generations.**

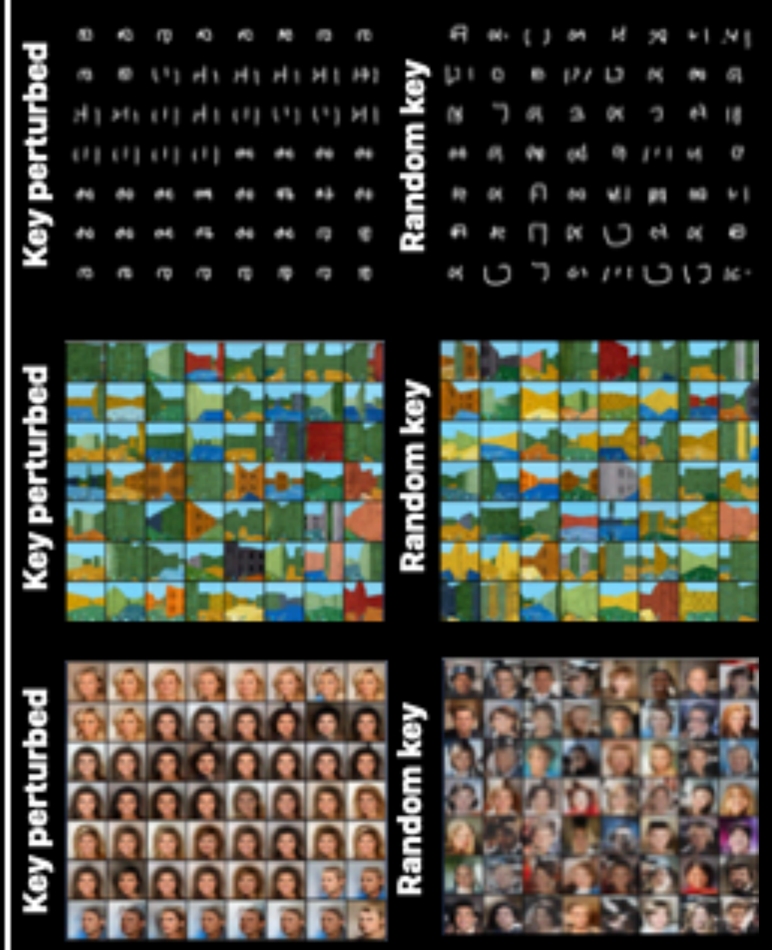


Model



	Binarized MNIST (nats/image)	Binarized Omniglot (nats/image)	Fashion MNIST (bits/dim)	CIFAR10 (bits/dim)	DMLab mazes (bits/dim)
VMA (Bornschein, 2017)	-	103.6	-	-	-
KM (Wu, 2018a)	-	≤ 68.3	-	≤ 4.37	-
DNC (Graves, 2016)	-	≤ 100	-	-	-
DKM (Wu 2018b)	≤ 75.3	≤ 77.2	-	≤ 4.79	≤ **2.75**
DKM w/ TSM (our impl)	≤51.84	≤70.88	≤4.15	≤ 4.31	≤ 2.92
K++ (ours)	≤**41.58**	≤**66.24**	≤**3.40**	≤ **3.28**	≤ 2.88

Generations



Denoising