

A New Era for
Associative Memories

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't
mttaer in waht oredr the Itteers in a wrod are, the olny
iprmoetnt tihng is taht the frist and lsat Itteer be at the rghit
pclae. The rset can be a toatl mses and you can sitll raed it
wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed
ervey Iteter by istlef, but the wrod as a wlohe.

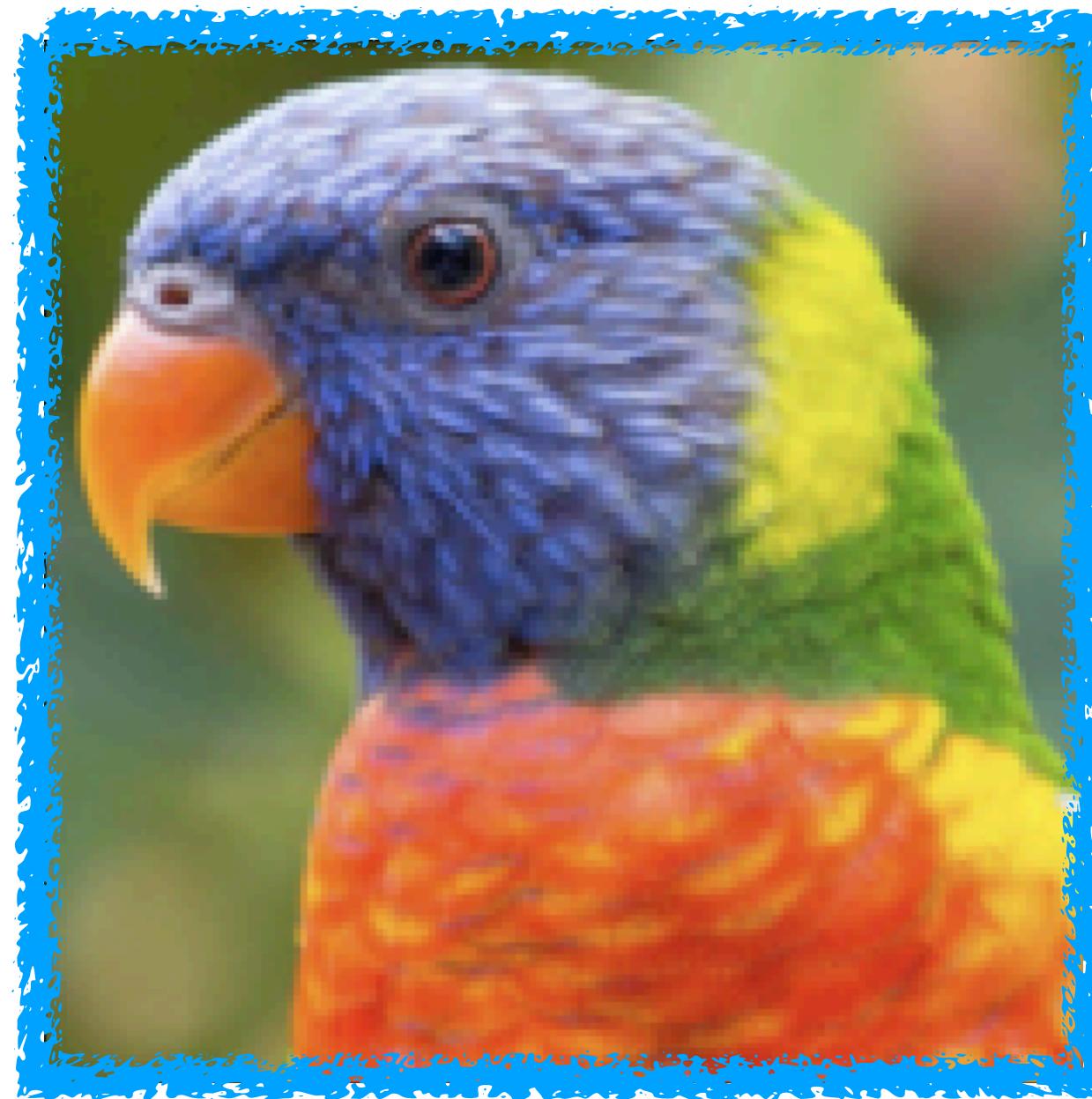
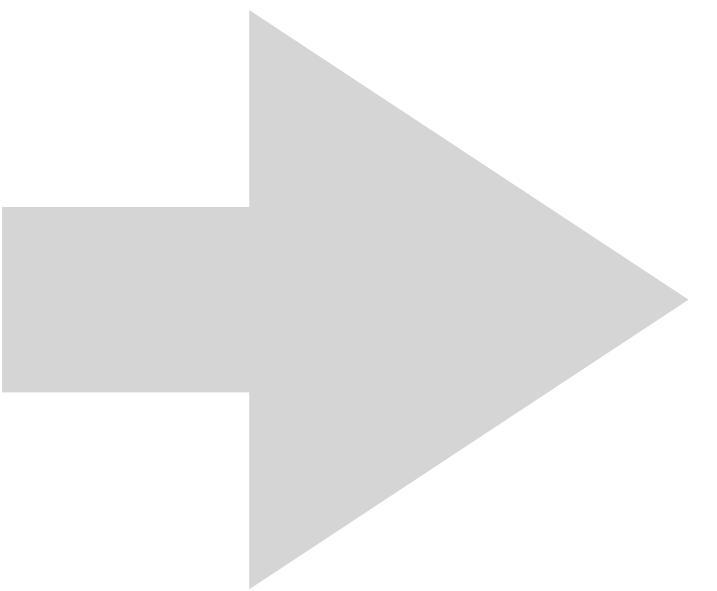
Modern AI is error correction

Self-supervision trains a model to **recreate data**
by **iteratively removing errors**

This is the M.O. of Associative Memory

Modern AI is error correction

Vision Transformer (ViT) : In-paint masked image tokens

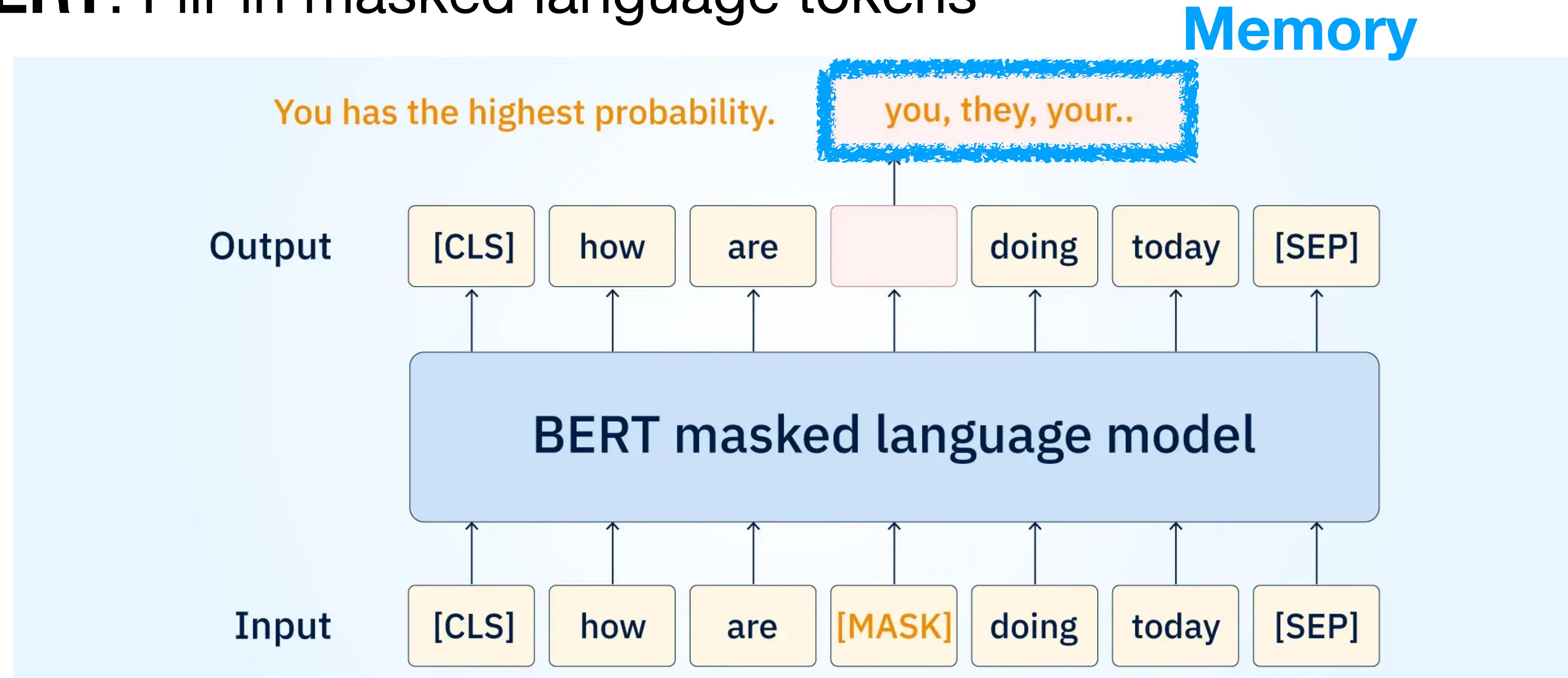


Memory

Training objective: minimize MSE between MASKed predictions and true pixels

Modern AI is error correction

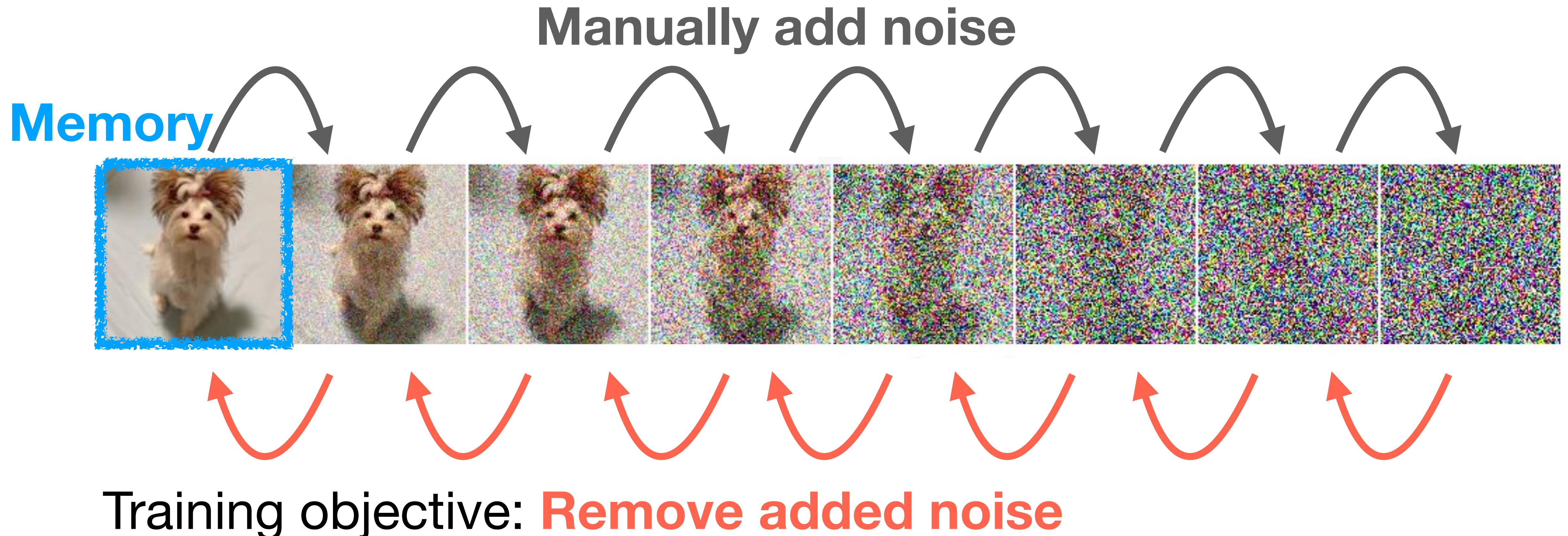
BERT: Fill-in masked language tokens



Training objective: minimize KL Div. between [MASK]ed predictions and true tokens

Modern AI is error correction

Diffusion Models: Recurrently remove noise from input



Modern AI is error correction

GPT: Predict the next token



Training objective: minimize KL Div. between next predicted and true token

Formal Associative Memory

- Fully Analog

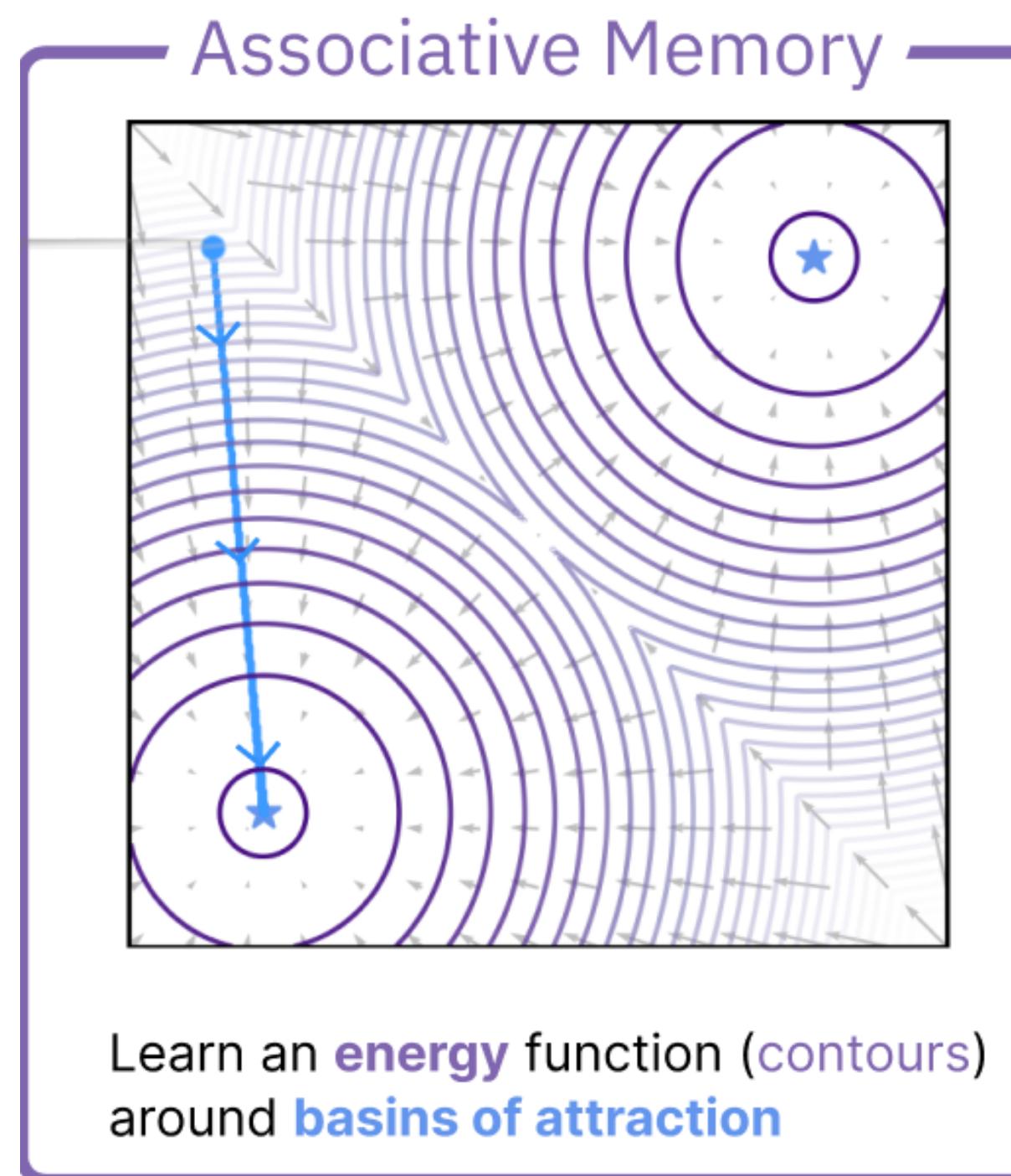
Computation is described in terms of a **dynamical system**.

$$\tau \frac{dx_{iA}}{dt} = -\frac{\partial E}{\partial g_{iA}}$$

Think beyond efficient matrix-multiplications. The whole model can be implemented on analog hardware

Advantages of Formal Associative Memory

- Fully Analog
- **Fixed Point Attractor**



Formal treatment of:

1. A **model-optimal** prediction
2. "**Thinking longer**" to solve harder problems
3. **Storage capacity**

Advantages of Formal Associative Memory

- Fully Analog
- Fixed Point Attractor
- **Simplicity**

Reformulating the Transformer
as Associative Memory

20x reduction in parameters

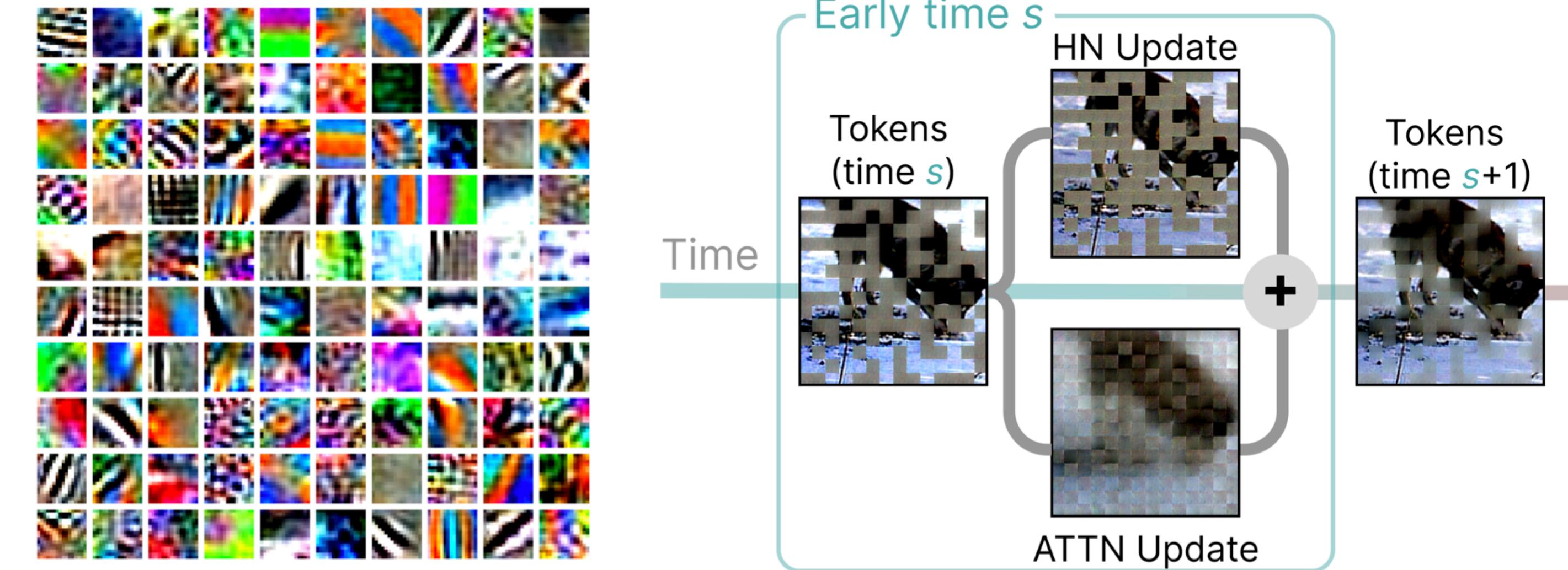
Parameter Efficiency

Model	NParams	NParams (per block)	
ViT-Base	86.28M ▼0.00%	7.08M	▼0.00%
ALBERT ViT-Base	8.41M ▲90.25%	7.08M	▼0.00%
ET	4.87M ▲94.36%	3.54M	▲50.02%

Advantages of Formal Associative Memory

- Fully Analog
- Fixed Point Attractor
- Simplicity
- Transparency

Interpretable



An input is **attracted** to an error-free version of itself stored in the network

Advantages of Formal Associative Memory

- Fully Analog
- Fixed Point Attractor
- Simplicity
- Transparency

Relatable

