

Holographic Reduced Representations

Plate's system, step by step

James Pustejovsky

Brandeis University
Department of Computer Science

CS 135: Computational Semantics

Fall 2026

Roadmap

I	Plate's ingredients	atoms; binding by convolution; frames
II	Taking a binding apart	the involution; correlation; the noise; unitary vectors
III	Frames at work	querying; cleanup; order and similarity
IV	How much fits	counting noise; capacity; nesting; sequences
V	The Fourier view, and the lineage	why the involution; where HRR came from; what comes next
VI	Close	summary; exercises; reading

Session 11c built circular convolution as the fold of the outer product, and proved that it cannot keep argument order. Today we build the whole system Plate (1995) put on top of it, computing by hand where the vectors are small and measuring where they are not.

Where We Are

Binding has one job: hold a role and its filler in one vector, so that which-is-which survives. Session 11c's route, in four steps:

outer product	$\mathbf{r}\mathbf{f}^\top$ keeps everything, at a cost of n^2 numbers per pair
the fold	sum it along wrapped diagonals, back down to n numbers
the result	circular convolution, $(\mathbf{r} \circledast \mathbf{f})_j = \sum_k r_k f_{(j-k) \bmod n}$
the price	with $\mathbf{u} = (1, 2, 0, 1)$, $\mathbf{v} = (2, 0, 1, 1)$: $\mathbf{u} \circledast \mathbf{v} = (4, 5, 2, 5) = \mathbf{v} \circledast \mathbf{u}$. The fold erases the antisymmetric half, so the binding cannot tell role from filler

11c ended with three questions it left open:

- 1 Why is the noise small? 11c's unbinding returned 0.599, against distractors near 0.
- 2 How do we get from a noisy vector back to a symbol?
- 3 How much fits in one vector?

Plate's Holographic Reduced Representations (HRR) answer all three. That is today.

Three Questions for Today

- ① **How do you take a binding apart?** Given the vector for $\text{agt} \circledast \text{mark}$ and the role agt , recover mark . Convolution has no obvious “division,” so what plays its part?
- ② **Why is the answer only approximately right, and how wrong is it?** We will find the error exactly at $n = 4$, then count it at $n = 1,024$.
- ③ **How much fits?** A sentence like *Hunger caused Mark to eat the fish* is several bindings inside one vector, one of them a whole clause. When do the answers stop coming back?

And a fourth, for the end: why Plate called these representations *holographic*.

Part I

Plate's Ingredients

Atoms: Random Vectors of Length About One

Every symbol gets an **atom**: each word (mark, fish), each role (agt, obj), each predicate (eat). Plate's rule: draw the n coordinates independently, with mean 0 and variance $1/n$. A normal distribution works, and so do the values $\pm 1/\sqrt{n}$.

n	length of an atom	dot product of two different atoms
64	1.00 ± 0.08	0.00 ± 0.13
1,024	1.00 ± 0.02	0.00 ± 0.03
16,384	1.00 ± 0.006	0.00 ± 0.008

Measured over 500 draws; “ $\pm$ ” is one standard deviation.

Two facts follow. **Every atom has length about 1**, so a dot product between atoms is already a cosine. **Two different atoms are nearly orthogonal**, within about $1/\sqrt{n}$: the fact Session 11c used to explain why the fold's collisions are survivable.

Nothing in an atom's coordinates says whether it is a word or a role; that lives in how the atoms are used. We write n for the dimension, as Plate does; Session 11c's tables and the next session write d .

Binding Is Circular Convolution

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

To bind **u** and **v**, output j adds every product $u_k v_m$ whose indices sum to j , wrapping around mod n :

$$(\mathbf{u} \circledast \mathbf{v})_j = \sum_{k=0}^{n-1} u_k v_{(j-k) \bmod n}$$

output	the four products
$j = 0$	$u_0 v_0 + u_1 v_3 + u_2 v_2 + u_3 v_1 = 1 \cdot 2 + 2 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 \quad 4$
$j = 1$	$u_0 v_1 + u_1 v_0 + u_2 v_3 + u_3 v_2 = 1 \cdot 0 + 2 \cdot 2 + 0 \cdot 1 + 1 \cdot 1 \quad 5$
$j = 2$	$u_0 v_2 + u_1 v_1 + u_2 v_0 + u_3 v_3 = 1 \cdot 1 + 2 \cdot 0 + 0 \cdot 2 + 1 \cdot 1 \quad 2$
$j = 3$	$u_0 v_3 + u_1 v_2 + u_2 v_1 + u_3 v_0 = 1 \cdot 1 + 2 \cdot 1 + 0 \cdot 0 + 1 \cdot 2 \quad 5$

So $\mathbf{u} \circledast \mathbf{v} = (4, 5, 2, 5)$: n numbers, not n^2 . In each row the indices add to j , so swapping **u** and **v** only reorders the products: $\mathbf{u} \circledast \mathbf{v} = \mathbf{v} \circledast \mathbf{u}$. Convolution is also associative, so $\mathbf{a} \circledast \mathbf{b} \circledast \mathbf{c}$ needs no brackets.

What a Binding Keeps, and What It Loses

Now with real atoms. Bind a role $\mathbf{c}$ to a filler $\mathbf{x}$ at $n = 1,024$ and measure:

property	measured	what it means
size	n numbers	same kind of object as an atom
unlike its inputs	$\cos(\mathbf{c} \circledast \mathbf{x}, \mathbf{c})$ and $\cos(\mathbf{c} \circledast \mathbf{x}, \mathbf{x})$: 0.00 ± 0.05	a new item
keeps similarity	$\cos(\mathbf{c} \circledast \mathbf{x}, \mathbf{c} \circledast \mathbf{y}) - \cos(\mathbf{x}, \mathbf{y})$: 0.00 ± 0.02	similar fillers, similar pairs
commutative	$\mathbf{c} \circledast \mathbf{x} = \mathbf{x} \circledast \mathbf{c}$, exactly	role and filler look alike

300 draws. For the similarity row, $\mathbf{y}$ was built with $\cos(\mathbf{x}, \mathbf{y}) = 0.6$.

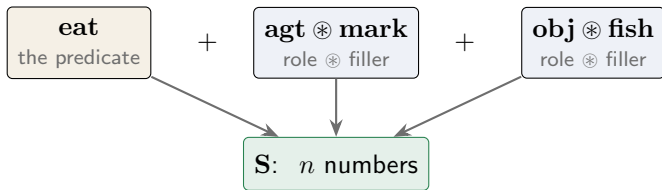
So $\text{agt} \circledast \text{mark}$ is unlike agt and unlike mark : a new item that means, roughly, *Mark as the agent*. And if mark and john were similar atoms, $\text{agt} \circledast \text{mark}$ and $\text{agt} \circledast \text{john}$ would be equally similar.

The last row is Session 11c's loss. Inside a binding, a role cannot be told from a filler. HRR therefore keeps roles in an inventory of their own, and argument order is carried by *which* role a filler is bound to.

Plate's Frames

A clause is a **frame**: its predicate, plus one binding per role, all added into one vector. Plate's own example (1995):

$$\text{Mark ate the fish} \mapsto \mathbf{S} = \text{eat} + \text{agt} \circledast \text{mark} + \text{obj} \circledast \text{fish}$$



Every term is a vector of length about 1, so $\mathbf{S}$ has length about $\sqrt{3}$, and it is *still* n numbers. A bigger frame adds more terms, never more dimensions.

Plate wrote predicate-specific roles, agt_{eat} and obj_{eat} ; we use one agt and one obj for all verbs, as in earlier sessions. Nothing below depends on the choice.

The question for Part II: given $\mathbf{S}$ and agt , how do we get **mark** back?

Part II

Taking a Binding Apart

What Would Undo a Convolution?

Convolution has an identity: the vector $\delta = (1, 0, \dots, 0)$. In the formula only $k = 0$ survives:

$$(\delta \circledast \mathbf{x})_j = \sum_k \delta_k x_{(j-k) \bmod n} = x_j, \quad \text{so} \quad \delta \circledast \mathbf{x} = \mathbf{x}$$

An **inverse** of $\mathbf{c}$ would be a vector $\mathbf{c}^{-1}$ with $\mathbf{c}^{-1} \circledast \mathbf{c} = \delta$. Then unbinding would be exact:

$$\mathbf{c}^{-1} \circledast (\mathbf{c} \circledast \mathbf{x}) = (\mathbf{c}^{-1} \circledast \mathbf{c}) \circledast \mathbf{x} = \delta \circledast \mathbf{x} = \mathbf{x}$$

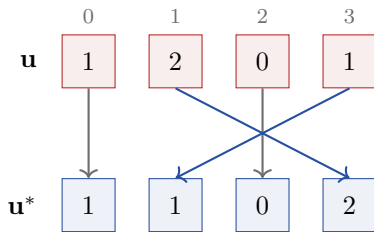
Most vectors have an inverse, but computing it means dividing, and Part V shows why dividing is dangerous once a vector holds several bindings.

Plate's choice is cheaper and more robust: an **approximate** inverse that needs no arithmetic at all, only a reordering of coordinates. It is called the **involution**.

The Involution, Step by Step

Keep the first coordinate and reverse the rest:

$$\mathbf{c}^* = (c_0, c_{n-1}, c_{n-2}, \dots, c_1), \quad \text{that is,} \quad c_i^* = c_{(-i) \bmod n}$$



For $\mathbf{u} = (1, 2, 0, 1)$: position 0 stays, positions 1 and 3 trade places, and position 2, its own mirror image mod 4, stays:

$$\mathbf{u}^* = (1, 1, 0, 2)$$

Do it twice and $\mathbf{u}$ comes back, $(\mathbf{u}^*)^* = \mathbf{u}$: that is what makes it an *involution*.

You have met this vector before. In 11c's fold, the first step reversed $\mathbf{v}$ and rotated it once, giving $(2, 1, 1, 0)$ – which is $\mathbf{v}^*$.

Correlation: Sliding One Vector Along Another

Plate unbinds by convolving with the involution. The operation has its own name, **circular correlation**, written $\#$:

$$\mathbf{c} \# \mathbf{t} = \mathbf{c}^* \circledast \mathbf{t}, \quad \text{written out:} \quad (\mathbf{c} \# \mathbf{t})_j = \sum_k c_k t_{(k+j) \bmod n}$$

Read the right-hand formula as a recipe: **shift $\mathbf{t}$ back by j places, then take its dot product with $\mathbf{c}$** . Entry j measures how well $\mathbf{c}$ matches $\mathbf{t}$ at offset j ; entry 0 is the plain dot product $\mathbf{c} \cdot \mathbf{t}$.

	convolution $\mathbf{c} \circledast \mathbf{t}$	correlation $\mathbf{c} \# \mathbf{t}$
indices in each product	add to j : $c_k t_{j-k}$	differ by j : $c_k t_{k+j}$
job in HRR	bind	unbind
order of arguments	does not matter	matters: $\mathbf{c} \# \mathbf{t} \neq \mathbf{t} \# \mathbf{c}$ in general

The name comes from signal processing, where correlating a known pattern against a signal finds where the pattern occurs.

Why Convolving With the Involution Slides

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

Write out $\mathbf{c}^* \circledast \mathbf{t}$ with the convolution formula, use $c_k^* = c_{-k}$, and rename $m = -k$ (all indices mod n):

$$(\mathbf{c}^* \circledast \mathbf{t})_j = \sum_k c_k^* t_{j-k} = \sum_k c_{-k} t_{j-k} = \sum_m c_m t_{j+m} = (\mathbf{c} \# \mathbf{t})_j$$

As k runs over every index, so does m . Check it with $\mathbf{u}^* = (1, 1, 0, 2)$ and $\mathbf{t} = \mathbf{u} \circledast \mathbf{v} = (4, 5, 2, 5)$:

j	convolve $\mathbf{u}^*$ with $\mathbf{t}$ (indices add to j)	slide $\mathbf{u}$ along $\mathbf{t}$ (indices differ by j)
0	$1 \cdot 4 + 1 \cdot 5 + 0 \cdot 2 + 2 \cdot 5 = 19$	$1 \cdot 4 + 2 \cdot 5 + 0 \cdot 2 + 1 \cdot 5 = 19$
1	$1 \cdot 5 + 1 \cdot 4 + 0 \cdot 5 + 2 \cdot 2 = 13$	$1 \cdot 5 + 2 \cdot 2 + 0 \cdot 5 + 1 \cdot 4 = 13$

The same four products appear in both columns, in a different order. The involution only re-indexes the role, so that convolution's "indices add to j " becomes correlation's "indices differ by j ." Plate gives $\#$ its own symbol, but it is always convolution with the involution.

Decoding by Hand

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

Bind, $\mathbf{t} = \mathbf{u} \circledast \mathbf{v} = (4, 5, 2, 5)$. To recover $\mathbf{v}$, correlate with $\mathbf{u}$: shift $\mathbf{t}$ back by j and dot with $\mathbf{u}$.

j	$\mathbf{t}$ shifted back by j , dotted with $\mathbf{u} = (1, 2, 0, 1)$	$(\mathbf{u} \# \mathbf{t})_j$
0	$(4, 5, 2, 5) \cdot \mathbf{u} = 4 + 10 + 0 + 5$	19
1	$(5, 2, 5, 4) \cdot \mathbf{u} = 5 + 4 + 0 + 4$	13
2	$(2, 5, 4, 5) \cdot \mathbf{u} = 2 + 10 + 0 + 5$	17
3	$(5, 4, 5, 2) \cdot \mathbf{u} = 5 + 8 + 0 + 2$	15

So $\mathbf{u} \# (\mathbf{u} \circledast \mathbf{v}) = (19, 13, 17, 15)$. Compare it with the candidates:

$$\cos((19, 13, 17, 15), \mathbf{v}) = 0.88, \quad \cos((19, 13, 17, 15), \mathbf{u}) = 0.76$$

Closer to $\mathbf{v}$ than to anything else here – but it is not $\mathbf{v}$. At $n = 4$ everything overlaps ($\mathbf{u}$ and $\mathbf{v}$ themselves sit at 0.50), so the margin is thin. The next slide finds the error exactly.

Where the Error Comes From

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

Convolution is commutative and associative, so the decoding can be regrouped:

$$\mathbf{u} \# (\mathbf{u} \circledast \mathbf{v}) = \mathbf{u}^* \circledast \mathbf{u} \circledast \mathbf{v} = (\mathbf{u} \# \mathbf{u}) \circledast \mathbf{v}$$

$\mathbf{u} \# \mathbf{u}$ is $\mathbf{u}$ correlated with itself, its **autocorrelation**. Entry j dots $\mathbf{u}$ with itself shifted by j :

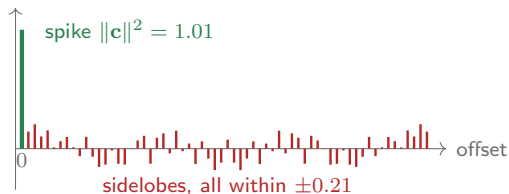
$$\mathbf{u} \# \mathbf{u} = (\underbrace{6}_{\mathbf{u} \cdot \mathbf{u}}, \underbrace{3, 4, 3}_{\text{sidelobes}}) = 6\boldsymbol{\delta} + (0, 3, 4, 3)$$

The **spike** at offset 0 is $\|\mathbf{u}\|^2$; the **sidelobes** are the overlaps at every other offset. Convolve with $\mathbf{v}$:

$$\mathbf{u} \# (\mathbf{u} \circledast \mathbf{v}) = \underbrace{6\mathbf{v}}_{(12, 0, 6, 6)} + \underbrace{(0, 3, 4, 3) \circledast \mathbf{v}}_{(7, 13, 11, 9)} = (19, 13, 17, 15) \checkmark$$

Decoding returns $\|\mathbf{u}\|^2 \mathbf{v}$ plus the sidelobes convolved with $\mathbf{v}$: **exact only when every sidelobe is zero.**

At Large n the Sidelobes Are Small, but Many



For a random atom at $n = 1,024$, measured over 300 atoms:

- spike: 1.00 ± 0.04
- each sidelobe: about ± 0.03 , that is $\pm 1/\sqrt{n}$
- but there are 1,023 of them, with total energy 1.00

the autocorrelation $c \# c$ of one random atom, $n = 64$

So the noise in $c \# (c \circledast x) = x + \text{noise}$ is about as long as x itself, and points in a random direction. Two equal lengths at right angles give

$$\cos(c \# (c \circledast x), x) \approx \frac{1}{\sqrt{1+1}} = 0.71 \quad \text{measured: } 0.71 \pm 0.02$$

Against any other atom the decoded vector scores 0.00 ± 0.03 . Recognizable, not exact: 0.71 against a floor of 0.03.

Unitary Vectors: Decoding Without Error

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

Some vectors have no sidelobes at all. Take $\mathbf{r} = \frac{1}{2}(1, 1, 1, -1)$: its autocorrelation is $\mathbf{r} \# \mathbf{r} = (1, 0, 0, 0) = \delta$. Bind $\mathbf{v}$ to it, then unbind with $\mathbf{r}^* = \frac{1}{2}(1, -1, 1, 1)$:

bind: $\mathbf{r} \circledast \mathbf{v}$			
0	$\frac{1}{2}(2 + 1 + 1 - 0)$	2	
1	$\frac{1}{2}(0 + 2 + 1 - 1)$	1	
2	$\frac{1}{2}(1 + 0 + 2 - 1)$	1	
3	$\frac{1}{2}(1 + 1 + 0 - 2)$	0	

unbind: $\mathbf{r}^* \circledast (2, 1, 1, 0)$			
0	$\frac{1}{2}(2 - 0 + 1 + 1)$	2	
1	$\frac{1}{2}(1 - 2 + 0 + 1)$	0	
2	$\frac{1}{2}(1 - 1 + 2 + 0)$	1	
3	$\frac{1}{2}(0 - 1 + 1 + 2)$	1	

$\mathbf{r} \# (\mathbf{r} \circledast \mathbf{v}) = (2, 0, 1, 1) = \mathbf{v}$, exactly. Each row is the convolution formula, $(\mathbf{a} \circledast \mathbf{b})_j = \sum_k a_k b_{(j-k) \bmod n}$, with $\mathbf{a} = \mathbf{r}$ in the first table and $\mathbf{a} = \mathbf{r}^*$ in the second.

Such vectors are called **unitary**. For them the involution *is* the exact inverse, $\mathbf{r}^* = \mathbf{r}^{-1}$. Part V gives the test: all of $\mathbf{r}$'s Fourier coefficients have size 1.

The Cyclic Shift Is a Unitary Binding

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

The simplest unitary vector is a unit impulse one place along, $\delta_1 = (0, 1, 0, 0)$. Convolution with it moves every coordinate one place right, and the last one to the front:

$$\delta_1 \circledast \mathbf{v} = (v_3, v_0, v_1, v_2) = (1, 2, 0, 1)$$

Its involution is $\delta_1^* = (0, 0, 0, 1)$, which shifts one place back, so the shift is undone exactly.

The cyclic shift ρ that the next session uses to mark word order is convolution with a unit impulse: a special HRR binding, unitary, and so exactly invertible. Shifting i places is binding with δ_i .

A small surprise in our running example: $\delta_1 \circledast \mathbf{v} = (1, 2, 0, 1) = \mathbf{u}$. Our $\mathbf{u}$ is $\mathbf{v}$ shifted by one place. Exercise 1 uses this.

Why Plate Used the Involution Anyway

An exact inverse exists for most atoms. Compare it with the involution on traces that hold k role-filler pairs, decoding one of them, at $n = 1,024$:

pairs in the trace, k	1	2	3	5	10
exact inverse $\mathbf{c}^{-1}$	1.00	0.34	0.25	0.18	0.12
involution $\mathbf{c}^*$	0.71	0.58	0.50	0.41	0.30

Mean cosine to the correct filler over 300 traces.

With one pair the exact inverse is perfect. Add a second pair and it falls below the involution, and stays there. The reason, in Part V: the exact inverse *divides* by the atom's Fourier coefficients, and a random atom always has a few small ones (the smallest is about 0.04 at this n). Whatever else sits in the trace at those frequencies is multiplied by up to about 28.

Plate notes that inverse filters are sensitive to noise, so on superposed traces the involution gives the better signal-to-noise ratio – and for unitary atoms the two coincide.

In-Class Exercise 1

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

- 1 Compute $\mathbf{v}^*$, then decode the other way: $\mathbf{v} \# (\mathbf{u} \circledast \mathbf{v})$. Which of $\mathbf{u}$, $\mathbf{v}$ is the result closest to?
- 2 Compute $\mathbf{v} \# \mathbf{v}$, the autocorrelation of $\mathbf{v}$. Compare it with $\mathbf{u} \# \mathbf{u} = (6, 3, 4, 3)$, and explain the match.
- 3 Show that $\frac{1}{2}(1, -1, 1, 1)$ is unitary by computing its autocorrelation. How is it related to $\mathbf{r} = \frac{1}{2}(1, 1, 1, -1)$?
- 4 ★ Decoding gave $6\mathbf{v}$ plus noise rather than $\mathbf{v}$ plus noise. Where does the 6 come from, and what does Plate's variance $1/n$ buy?

Seven minutes. Item 4 is the starred one.

In-Class Exercise 1: Answers

1. $\mathbf{v}^* = (2, 1, 1, 0)$. Shifting $\mathbf{t} = (4, 5, 2, 5)$ back by j and dotting with $\mathbf{v}$ gives $\mathbf{v} \# \mathbf{t} = (15, 19, 13, 17)$, at cosine 0.88 with $\mathbf{u}$. Either factor of a binding decodes the other, because the binding is commutative.
2. $\mathbf{v} \# \mathbf{v} = (6, 3, 4, 3)$, the same as $\mathbf{u}$'s. Since $\mathbf{u} = \delta_1 \circledast \mathbf{v}$ is $\mathbf{v}$ shifted, every offset compares the same pairs of numbers. A shift never changes an autocorrelation.
3. Its autocorrelation is $\frac{1}{4}(4, 0, 0, 0) = \delta$: offset 0 gives $\frac{1}{4}(1 + 1 + 1 + 1)$, offset 1 gives $\frac{1}{4}(-1 - 1 + 1 + 1)$, offset 2 gives $\frac{1}{4}(1 - 1 + 1 - 1)$, and offset 3 matches offset 1, since an autocorrelation is symmetric. It is $\mathbf{r}^*$: the involution of a unitary vector is unitary.
4. ★ The spike is $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2 = 6$. With variance $1/n$ every atom has $\|\mathbf{c}\|^2 \approx 1$, so decoding returns the filler at its own scale, cleanup can compare it directly with stored atoms, and the terms of a frame stay equally weighted.

Part III

Frames at Work

Querying a Frame, Term by Term

Ask *Mark ate the fish* who did the eating: correlate the frame with **agt**. Correlation distributes over the sum, term by term:

$$\mathbf{agt} \# \mathbf{S} = \underbrace{\mathbf{agt} \# \mathbf{eat}}_{\text{noise}} + \underbrace{\mathbf{agt} \# (\mathbf{agt} \circledast \mathbf{mark})}_{\text{mark} + \text{sidelobe noise}} + \underbrace{\mathbf{agt} \# (\mathbf{obj} \circledast \mathbf{fish})}_{\text{noise}}$$

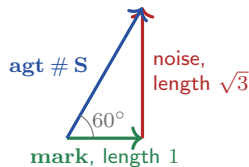
Count the pieces. One copy of **mark**, of length about 1.

Three noise pieces, each of length about 1 and pointing in unrelated directions, so together of length about $\sqrt{3}$, at right angles to **mark**:

$$\cos(\mathbf{agt} \# \mathbf{S}, \mathbf{mark}) \approx \frac{1}{\sqrt{1+3}} = 0.50$$

Measured: 0.50, over 400 frames at $n = 1,024$.

The noise is only nearly perpendicular to **mark**, as any random direction is at large n . Total length 2, so $\cos 60^\circ = 1/2$.



Cleanup Returns the Filler

The decoded vector is not an atom. A **cleanup memory** holds every atom we minted and returns the one with the highest cosine. One frame, $n = 1,024$, nine atoms in the memory:

question	compute	best atom	runner-up
who ate?	agt # S	mark 0.47	0.04
what was eaten?	obj # S	fish 0.49	0.03
what role does the fish fill?	fish # S	obj 0.49	0.06
what role does Mark fill?	mark # S	agt 0.46	0.06

Atoms in the memory: eat, see, cause, mark, john, fish, hunger, agt, obj. Over 400 frames the right atom averages 0.50; every other atom, 0.00 ± 0.03 .

Unbinding works in both directions: by a role, to get its filler, or by a filler, to get its role, because the binding is commutative.

Cleanup is where the symbol comes back. Everything before it is continuous; cleanup rounds to a finite list. The next session returns to what that rounding means.

Order Lives in the Roles

Compare *Mark ate the fish*, $S_1 = \text{eat} + \text{agt} \circledast \text{mark} + \text{obj} \circledast \text{fish}$, with other frames built the same way:

frame	terms shared with S_1	$\cos(S_1, \cdot)$	bag of words
<i>The fish ate Mark</i>	eat	0.33	1.00
<i>John ate the fish</i>	eat, $\text{obj} \circledast \text{fish}$	0.67	0.67
<i>Mark saw the fish</i>	$\text{agt} \circledast \text{mark}$, $\text{obj} \circledast \text{fish}$	0.67	0.67
<i>John saw Mark</i>	none	0.00	0.33

Mean over 400 draws at $n = 1,024$. Each frame has three terms of length about 1, so sharing s of them gives a cosine of about $s/3$.

HRR similarity counts shared role–filler pairs, not shared words. *The fish ate Mark* has every word of *Mark ate the fish*, so a bag of words scores them as identical; HRR keeps only the predicate in common, because mark and fish have traded roles.

This is Session 11c's diagnosis, answered the HRR way: the binding cannot see order, but the choice of role can.

Part IV

How Much Fits

Counting the Noise

Decode one pair from a trace that holds k pairs. The signal is one filler, of length 1. The noise has k pieces of length about 1: the other $k - 1$ pairs, plus the decoding role's own sidelobes. So

$$\cos \approx \frac{1}{\sqrt{k+1}} \quad \text{and, with unitary roles (no sidelobes),} \quad \cos \approx \frac{1}{\sqrt{k}}$$

pairs in the trace, k	1	2	3	5	9	20
random roles, measured	0.71	0.58	0.50	0.41	0.31	0.22
$1/\sqrt{k+1}$	0.71	0.58	0.50	0.41	0.32	0.22
unitary roles, measured	1.00	0.71	0.58	0.45	0.33	0.22
$1/\sqrt{k}$	1.00	0.71	0.58	0.45	0.33	0.22

$n = 1,024$, mean of 200 traces. A frame's predicate counts as one more term, which is why *Mark ate the fish* gave 0.50: three terms, so $k = 3$.

Unitary roles help most when the trace is small; by $k = 20$ the other pairs swamp the difference.

Session 11c's *John loves Mary* trace held $k = 2$ pairs: predicted $1/\sqrt{3} = 0.58$, and 11c measured 0.599 in one trial.

When Cleanup Fails

Cleanup fails when some wrong atom outscores the right one. Each wrong atom scores about $\pm 1/\sqrt{n}$, and the best of m wrong atoms reaches about

$$\frac{\sqrt{2 \ln m}}{\sqrt{n}} \quad \text{for } m = 1,000 \text{ and } n = 1,024: \quad \frac{\sqrt{13.8}}{32} \approx 0.12$$

The right atom scores about $1/\sqrt{k+1}$. Trouble starts when the two meet:

$$\frac{1}{\sqrt{k+1}} \approx \frac{\sqrt{2 \ln m}}{\sqrt{n}} \iff k+1 \approx \frac{n}{2 \ln m} \implies k \approx 70 \text{ at } n = 1,024$$

The dimension needed grows linearly with the number of pairs, and only logarithmically with the size of the cleanup memory. Plate (1995) derives the same shape with care; this back-of-the-envelope version gets the scaling right.

The $\sqrt{2 \ln m}$ is the usual size of the largest of m independent standard normal scores. Doubling the vocabulary barely moves it; doubling the number of pairs needs twice the dimension.

Measured Capacity

Bind k random role-filler pairs into one trace, decode the first role, and clean up against $m = 1,000$ fillers. Fraction of trials returning the right filler:

n	$k = 1$	2	5	10	20	50	100
256	1.00	1.00	1.00	0.93	0.61	0.18	0.05
1,024	1.00	1.00	1.00	1.00	1.00	0.85	0.46
4,096	1.00	1.00	1.00	1.00	1.00	1.00	1.00

150 trials per cell.

At $n = 1,024$ the first failures appear between $k = 20$ and $k = 50$, and half the trials fail near $k = 100$; the estimate of about 70 from the last slide falls in between. **Quadrupling n moves the breaking point about four times further out:** at $n = 256$, half the trials fail near $k = 25$.

The failure is a cliff, not a slope. Size n for the largest structure you intend to store.

Nesting: *Hunger Caused Mark to Eat the Fish*

Plate's recursive example. A whole frame can fill a role, because it is just another vector of n numbers:

$$\mathbf{E} = \text{eat} + \text{agt} \circledast \text{mark} + \text{obj} \circledast \text{fish}, \quad \mathbf{C} = \text{cause} + \text{agt} \circledast \text{hunger} + \text{obj} \circledast \mathbf{E}$$

Who ate? needs two decodings: first $\text{obj} \# \mathbf{C}$, to reach the caused event, then agt on the result.
Measured at $n = 1,024$:

step	compared with	cos
$\text{obj} \# \mathbf{C}$	$\mathbf{E}$	0.67
$\text{agt} \# (\text{obj} \# \mathbf{C})$, no cleanup in between	mark	0.26
clean up $\text{obj} \# \mathbf{C}$ to the stored $\mathbf{E}$, then decode with agt	mark	0.50

Cleanup chose $\mathbf{E}$ over three other stored frames in all 400 trials.

Clean up at every level. The noise of two decodings compounds; cleanup in between resets it. That requires $\mathbf{E}$ itself to sit in cleanup memory, as one item. Hinton (1990) called such a stored, compressed whole a **reduced description**, and it is the “reduced” in Plate’s name.

Sequences, and the Shift Again

Order within a string can be carried by powers of one unitary key $\mathbf{p}$, written $\mathbf{p}^2 = \mathbf{p} \circledast \mathbf{p}$ and so on:

$$\textit{the dog barked} \mapsto \mathbf{p} \circledast \textit{the} + \mathbf{p}^2 \circledast \textit{dog} + \mathbf{p}^3 \circledast \textit{barked}$$

To read position 2, correlate with $\mathbf{p}^2$. Powers of a unitary vector are unitary, and different powers are nearly orthogonal, so each position decodes as one term among three. Measured at $n = 1,024$:

query	cos to the	cos to dog	cos to barked
$\mathbf{p} \# \mathbf{s}$	0.58	0.00	0.00
$\mathbf{p}^2 \# \mathbf{s}$	0.00	0.58	0.00
$\mathbf{p}^3 \# \mathbf{s}$	0.00	0.00	0.58

$0.58 = 1/\sqrt{3}$. Off-diagonal entries are 0.00 ± 0.03 .

Choose $\mathbf{p} = \delta_1 = (0, 1, 0, \dots, 0)$ and the powers are shifts: $\delta_1^i \circledast \mathbf{x}$ moves $\mathbf{x}$ by i places. That is how the next session encodes word order, writing $\rho^i(\mathbf{x})$ for a shift by i places.

In-Class Exercise 2

- 1 A frame has a predicate and eight role bindings. Estimate the cosine between the correct filler and the vector you get by decoding one role.
- 2 Is cleanup safe for that frame at $n = 1,024$ with 1,000 atoms in memory? At $n = 256$?
- 3 With $\mathbf{C} = \text{cause} + \text{agt} \circledast \text{hunger} + \text{obj} \circledast \mathbf{E}$ for *Hunger caused Mark to eat the fish*, write the queries that answer *What caused it?* and *What was eaten?* Say where cleanup is needed.
- 4 ★ Nested queries worked only because $\mathbf{E}$ was stored in cleanup memory. What does that say about where the symbols live in HRR?

Seven minutes. Item 4 is the starred one.

In-Class Exercise 2: Answers

1. Nine terms, so the decoded vector has one copy of the filler and nine noise pieces:
 $\cos \approx 1/\sqrt{9+1} \approx 0.32$. Measured: 0.31.
2. The best wrong atom reaches about $\sqrt{2 \ln 1000}/\sqrt{n}$: 0.12 at $n = 1,024$, comfortably below 0.32. At $n = 256$ it is about 0.23, **too close for comfort**; the capacity table shows about 7% failures at $k = 10$.
3. *What caused it?*: clean up **agt** # **C**, which returns hunger. *What was eaten?*: clean up **obj** # **C** against stored frames, which returns **E**; then clean up **obj** # **E**, which returns fish. **Two levels, two cleanups.**
4. ★ In the cleanup memory: the list of atoms and of stored frames, which the next session calls the *codebook*. **The vectors and the operations are continuous; the system becomes symbolic each time it rounds to that list.** The next session ends on this claim.

Part V

The Fourier View, and the Lineage

The Fourier View

index	0	1	2	3
u	1	2	0	1
v	2	0	1	1

The discrete Fourier transform of Session 11f, $(\mathcal{F}\mathbf{x})_k = \sum_j x_j e^{-2\pi i j k / n}$, turns convolution into coordinate-by-coordinate multiplication:

$$\mathcal{F}(\mathbf{u} \circledast \mathbf{v}) = \mathcal{F}\mathbf{u} \odot \mathcal{F}\mathbf{v} : \quad (4, 1-i, -2, 1+i) \odot (4, 1+i, 2, 1-i) = (16, 2, -4, 2)$$

and transforming back gives $(4, 5, 2, 5)$, as before. The involution *conjugates* the spectrum, $\mathcal{F}(\mathbf{u}^*) = \overline{\mathcal{F}\mathbf{u}}$, so decoding multiplies by the squared magnitudes:

$$\mathcal{F}(\mathbf{u} \# (\mathbf{u} \circledast \mathbf{v})) = |\mathcal{F}\mathbf{u}|^2 \odot \mathcal{F}\mathbf{v}, \quad |\mathcal{F}\mathbf{u}|^2 = (16, 2, 4, 2)$$

and $(16, 2, 4, 2)$ transforms back to $(6, 3, 4, 3)$, the autocorrelation of Part II.

Decoding is exact exactly when $|\mathcal{F}\mathbf{u}|^2$ is constant: a flat spectrum. That is the definition of **unitary**, every $|\mathcal{F}_k \mathbf{u}| = 1$. For $\mathbf{r} = \frac{1}{2}(1, 1, 1, -1)$, $\mathcal{F}\mathbf{r} = (1, -i, 1, i)$.

In practice HRR is computed this way: two FFTs, one coordinate-wise product, one inverse FFT, at a cost of $O(n \log n)$.

The Exact Inverse Divides

The exact inverse has spectrum $1/\mathcal{F}\mathbf{c}$: it divides each frequency by the atom's coefficient there. Take $\mathbf{c} = (1, 1, 1, 0.9)$:

$$|\mathcal{F}\mathbf{c}| = (3.9, 0.1, 0.1, 0.1) \implies 1/|\mathcal{F}\mathbf{c}| = (0.26, 10, 10, 10)$$

and transforming back gives $\mathbf{c}^{-1} = (2.56, -7.44, 2.56, 2.56)$, more than four times as long as $\mathbf{c}$.

	exact inverse $\mathbf{c}^{-1}$	involution $\mathbf{c}^*$
multiplies frequency k by	$1/\mathcal{F}_k\mathbf{c}$	$\overline{\mathcal{F}_k\mathbf{c}}$
on this $\mathbf{c}$, frequencies 1–3	multiplied by 10	multiplied by 0.1
on a trace with one pair	exact	distorted where $\mathbf{c}$ is weak
on a trace with several pairs	amplifies the other pairs where $\mathbf{c}$ is weak	no division, so no blow-up

That is what we measured at $n = 1,024$: with one pair, 1.00 for the exact inverse against 0.71 for the involution; with two pairs, 0.34 against 0.58. There a random atom's smallest coefficient is typically about 0.04, a gain of about 28 on the other pairs. [The involution trades exactness on clean traces for safety on crowded ones](#); unitary atoms need no trade.

Why “Holographic,” and Where HRR Came From

1960s–70s	holographic and correlation memories: store a pair by convolving, recall by correlating, by analogy with optical holograms
1982	psychological models of memory for pairs built on convolution and correlation: Murdock's TODAM and Metcalfe Eich's CHARM
1990	Hinton's <i>reduced descriptions</i> : a whole structure compressed into a vector that fits one slot; Smolensky's tensor product binding
1991–95	Plate's HRR: circular convolution at fixed n , the involution, cleanup memory, a capacity analysis
2003	Plate's book, <i>Holographic Reduced Representation</i>

“Holographic” because convolution and correlation are the operations of those holographic memories, and, as in an optical hologram, every item is spread over every coordinate: damage to some coordinates lowers every cosine a little rather than erasing any one item. “Reduced” because a whole frame becomes one vector that can fill a slot of another frame, as in *Hunger caused Mark to eat the fish*.

So convolution memory was a model of human recall before Plate made it a representation for structure.

HRR Among Its Neighbors

	TPR (Smolensky)	HRR (Plate)	MAP (Gayler)
atoms	any vectors	random, variance $1/n$	random ± 1
binding	outer product $\mathbf{r}\mathbf{f}^\top$	circular convolution	coordinate-wise product
size of a pair	n^2	n	n
unbinding	exact, with orthonormal roles	approximate; exact for unitary roles	exact: binding is its own inverse
commutative?	no	yes	yes
cost per pair	n^2	$n \log n$, by FFT	n

Next session: MAP. With coordinates ± 1 , multiplying a vector by itself coordinate-wise gives all ones, so every atom undoes itself exactly. Unbinding becomes exact and computable by hand, with no Fourier analysis.

That frees the next session for the architecture as a whole: the three operations, cleanup, records, sequences, analogy without training, and types. Everything there has an HRR counterpart from today.

Summary

- 1 HRR atoms are random vectors with variance $1/n$: length about 1, and nearly orthogonal to each other.
- 2 **Binding** is circular convolution: n numbers, similar to neither input, similarity-preserving, and commutative. Order is carried by which role a filler is bound to.
- 3 **Unbinding** is correlation with the role, $\mathbf{c} \# \mathbf{t} = \mathbf{c}^* \circledast \mathbf{t}$, where the involution keeps the first coordinate and reverses the rest.
- 4 Decoding returns the filler plus the role's **sidelobes** convolved with it. At large n the sidelobes are small but many, so one pair decodes at about 0.71; unitary roles have none and decode exactly.
- 5 With k pairs the right answer scores about $1/\sqrt{k+1}$, against a best wrong answer near $\sqrt{2 \ln m}/\sqrt{n}$. Capacity grows linearly in n and fails as a cliff.
- 6 **Cleanup** rounds to the stored atoms and frames. Nested structures need it at every level.
- 7 The exact inverse divides by Fourier coefficients and amplifies the other pairs, so Plate decoded with the involution.

Exercises

- 1 By hand at $n = 4$: superpose the two pairs $\mathbf{r} \circledast \mathbf{v}$ and $\mathbf{u} \circledast \mathbf{v}$ into one trace, with $\mathbf{r} = \frac{1}{2}(1, 1, 1, -1)$, and decode with $\mathbf{r}$. Show that $\mathbf{r}$'s sidelobes contribute nothing, and compute what the other pair contributes.
- 2 In numpy, implement $\circledast$ and $\#$ with `np.fft`, using `np.roll(c[:-1], 1)` for the involution. Reproduce the 0.71 of Part II at $n = 1,024$.
- 3 Build *Mark ate the fish*, *The fish ate Mark*, and *John saw Mark*. Reproduce the similarity table of Part III, then make john and mark similar atoms ($\cos = 0.6$) and report what changes.
- 4 Reproduce the capacity table for $n = 256$ and $n = 1,024$, and plot accuracy against k/n . Where is the cliff, as a fraction of n ?
- 5 Make every role unitary by setting each Fourier coefficient's magnitude to 1. Measure the gain for $k = 1, 3, 10$, and compare with $1/\sqrt{k}$.
- 6 ★ Encode *Hunger caused Mark to eat the fish*. Query *Who ate?* with and without cleanup of the intermediate frame, then add a third level of nesting and report the cosine at each depth.

Reading

- T. A. Plate, Holographic reduced representations, *IEEE Transactions on Neural Networks* 6(3), 1995. The source for today: atoms, the involution, frames, capacity.
- T. A. Plate, *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*, CSLI Publications, 2003.
- G. E. Hinton, Mapping part-whole hierarchies into connectionist networks, *Artificial Intelligence* 46, 1990. Reduced descriptions.
- P. Smolensky, Tensor product variable binding and the representation of symbolic structures in connectionist systems, *Artificial Intelligence* 46, 1990.
- B. B. Murdock, A theory for the storage and retrieval of item and associative information, *Psychological Review* 89, 1982.
- J. Metcalfe Eich, A composite holographic associative recall model, *Psychological Review* 89, 1982.
- P. Kanerva, Hyperdimensional computing, *Cognitive Computation* 1, 2009. The wider family, for next session.