Non-Serial Inter-Area Synergy in the Macaque Ventral Stream

Arun Ram Ponnambalam*
Department of Biomedical Engineering
SRM Institute of Science and Technology

Krishnan Venkiteswaran Pottore*
Department of Computer Science
University College Dublin

Abstract

The ventral visual stream is widely modeled as a serial feedforward hierarchy in which V1, V4, and IT population codes develop sequentially during object recognition. We ask whether a second, concurrent coding mode exists—one organized not by anatomical order but by joint population structure across areas. Using Partial Information Decomposition applied to simultaneous multielectrode spiking recordings across all three areas at millisecond resolution—in two macaque monkeys viewing 25,000+ natural-image trials, we decompose population coding into serial (unique per area) and synergistic (joint across areas) components at 15 ms windows, 5 ms stride across four distinct ResNet-18 target representations spanning low-level spatial features to high-level object identity. Three findings are observed in both animals. First, synergistic inter-area coupling emerges before IT carries any unique object-related information—a dissociation of 15–35 ms that replicates in direction in every condition where IT-unique information is detected—such that the joint population integrates before the apex encodes; moreover, V1–IT synergy persists after V1's unique information reaches zero. Second, although V1↔IT and V1↔V4 coupling emerge simultaneously and rise in parallel, V1↔IT exhibits stronger peak synergy at mid-to-high-level targets in both animals, suggesting a prominent role for non-serial joint coding. Third, when V1 and V4 are treated as an integrated block, their synergy with IT emerges last across all tested conditions; this ordering is also expected under a feedforward account, so we do not infer a mechanism from it. Together, these results show that serial and synergistic population codes co-occur in the same recordings, overlap in time, but follow different organizational principles, providing a new level of nuance in our understanding of the primate ventral stream and introducing constraints for biologically grounded models of vision.

1 Introduction

The primate ventral visual stream maps a continuously varying retinal image onto a stable, transformation-tolerant representation of object identity within roughly 200 ms. The canonical account holds that this is achieved through a serial feedforward hierarchy in which V1 encodes oriented edges and spatial frequencies, V4 integrates mid-level features such as curvature, colour, and texture, and inferotemporal cortex (IT) arrives at a transformation-tolerant representation of object identity [1]. This cascade has been operationalised by deep CNNs trained on image classification: task-optimised hierarchical networks predict single-unit responses in macaque V4 and IT [3], earlier layers best predict V1 and V2 while deeper layers best predict V4 and downstream regions [5], and match quality increases with model depth and categorisation performance [2]. Deep supervised networks outperform earlier model classes —including HMAX, VisNet, and hand-crafted

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

feature models—in explaining IT representational geometry [4] and predicting individual IT spiking responses [3].

A strictly serial cascade carries specific, testable predictions: (i) representational content should emerge in the order V1→V4→IT; (ii) inter-area coupling should respect anatomical adjacency, with adjacent pairs (V1–V4, V4–IT) interacting more tightly than the non-adjacent pair (V1–IT); and (iii) each area's unique content should emerge before, not after, it begins to interact with other areas. Feedforward networks are not synergy-free under our measure. With MMI redundancy, synergy reduces to the gain in joint information over the best single source, $\mathrm{Syn} = I(\mathbf{Z}; \mathbf{Y}) - \max_r I(\mathbf{S}_r; \mathbf{Y})$, and we estimate information that is linearly accessible from a finite number of principal components. In a feedforward network, nonlinear operations between layers create features in a deeper layer that cannot be linearly read out from an earlier one, while the earlier layer retains detail that the deeper layer discards, so a linear readout of both layers together can carry more about a target than either alone. Consistent with this, we find nonzero synergy between non-adjacent layers in every feedforward architecture we test (Section 3.7). Such synergy can therefore arise from layered nonlinear computation read out linearly, and does not by itself indicate non-feedforward interaction.

Evidence already shows purely feedforward models fail in specific domains. Anatomical mapping reveals inter-areal projection weights spanning five orders of magnitude [11]; pharmacological and optogenetic inactivation of feedback alters response gain, surround suppression, and tuning in downstream regions [10]; and ConvRNNs incorporating recurrence and long-range feedback better match the temporal dynamics of V4 and IT than feedforward-only architectures [7].

Partial Information Decomposition (PID) [9] provides a suitable framework, decomposing total mutual information into non-negative atoms of unique, redundant, and synergistic information. Large-scale decomposition of human fMRI reveals redundant interactions in sensorimotor cortex and synergistic interactions in higher-order association cortex, with synergy exhibiting the greatest evolutionary expansion from macaque to human [12]. Extended to invasive electrophysiology, inter-areal ECoG in awake marmosets is synergistic across the auditory hierarchy, requiring strong long-range feedback in a brain-constrained model [13]; and in macaque and human visual cortex, broadband V1–V4 signals are highly synergistic while narrowband gamma is predominantly redundant—dissociating the information-theoretic roles of two concurrently active signal types [14].

The THINGS Ventral Stream Spiking Dataset (TVSD) [15]—chronic multi-unit recordings spanning V1, V4, and IT in two macaque monkeys viewing >25,000 natural images from the THINGS database [18]—provides the stimulus diversity, temporal resolution, and simultaneous multi-area coverage required to address this question. Population decoding establishes that small IT ensembles carry surprisingly accurate object identity information [16], but decodability from a single area is fundamentally different from asking what the circuit computes collectively. Correlations within and across areas structure both encoding and downstream transmission [17], and prior TVSD characterisations show noise correlations are strongest between similarly tuned neurons [15]—consistent with shared feedforward drive—but whether joint activity carries information beyond individual areas, and when, has, to our knowledge, not been reported.

Here we apply three-source PID to simultaneous spiking recordings from V1, V4, and IT in both animals viewing >22,000 unique natural images (25,248 trials), decomposing unique, redundant, and synergistic information in 15 ms windows stepped by 5 ms across the 300 ms trial for four distinct ResNet-18 targets spanning early convolutional to final identity layers. We report three findings, each replicating in both animals. First, inter-area synergy emerges 15–70 ms before IT carries any unique object-related information: IT participates in a joint code before it contributes as an independent encoder. Second, pairwise synergy emerges near-concurrently across area pairs, and at mid-to-high-level targets the non-adjacent V1–IT pair matches or exceeds the adjacent pairs, indicating a non-serial mode of inter-areal coding. Third, the V1+V4 block couples with IT later than every other block–area combination (167.5 ms in both animals at AvgPool). Because a feedforward account also predicts a late block–IT coupling, we read it only alongside the pairwise result, in which V1–IT coupling precedes block–IT coupling. Together these findings indicate a ventral stream in which a serial representational hierarchy, with unique information emerging in the V1→V4→IT sequence, coexists with a distributed synergistic code in which non-adjacent interactions are prominent. With DINO ViT-S/16 and CorNet-S as targets, V1–IT again carries the largest pairwise synergy at most targets (Section 3.6), whereas among the V1, V4, and IT areas of CorNet-S it is not the dominant pair (Section 3.7).

2 Methods

2.1 Dataset and Preprocessing

We analysed the THINGS Ventral Stream Spiking Dataset (TVSD) [15], which provides simultaneous high-channel-count multi-unit activity (MUA) recordings from macaque V1, V4, and IT during passive viewing of 25,248 natural images from the THINGS database [18] (25,248 trials: 22,248 training images shown once plus 100 test images shown 30 times). Each trial was recorded over a 300 ms window at 1 ms resolution, with stimulus onset at 100 ms, yielding a 100 ms pre-stimulus baseline and a 200 ms post-stimulus analysis window. Data from two animals are analysed independently; main-text findings are examined in both. Monkey F: 512 V1, 320 IT, 192 V4 channels. Monkey N: 960 channels after exclusion of one unstable array (Array 6, 64 channels): 448 V1, 256 V4, 256 IT. All remaining channels were retained. Each channel was z-scored across all time bins and trials jointly to equalise firing-rate scale across electrodes and animals. Neural population activity was summarised in non-overlapping 15 ms windows stepped by 5 ms, yielding 57 time bins with centres spanning 7.5–292.5 ms relative to stimulus onset. For each time bin $[t,\, t + W)$ with $W = 15$ ms, the neural observation for trial k and ROI r was obtained by averaging the z-scored MUA across the window.

2.2 CNN Representational Targets

We extracted population-code targets from five stages of a pre-trained ResNet-18 [19]: outputs of residual blocks `layer1`–`layer4` (global-average-pooled to 64, 128, 256, 512 dimensions respectively) and the final `avgpool` layer (512-d). Each was compressed to 50 PCs via standard PCA, yielding a target matrix $\mathbf{Y}^{(\ell)} \in \mathbb{R}^{N\times 50}$ per layer. The pooled `layer4` output and the `avgpool` output are identical, so "Layer 4" and "AvgPool" results are duplicates and there are four distinct targets. For DINO ViT-S/16 [34] (self-supervised) we used the mean patch-token representations (384-d) of blocks 3 and 9 in Monkey N. For CorNet-S [35] we used the spatially pooled final-timestep outputs of its V1, V4, and IT areas in Monkey F, which the model's authors define to correspond to those areas.

2.3 Marchenko–Pastur PCA for Signal Subspace Estimation

For each ROI r and time bin t, column-centred neural data $\mathbf{X}_r(t) \in \mathbb{R}^{N\times C_r}$ were projected onto a signal subspace estimated via the Marchenko–Pastur law [20, 21]. Noise variance was estimated as the median positive eigenvalue of the sample covariance $\hat{\mathbf{\Sigma}}_r = \frac{1}{N-1}\mathbf{X}_r^\top \mathbf{X}_r$,

$$\hat{\sigma}^2 = \mathrm{median}\{\lambda_i : \lambda_i > 0\}, \qquad \hat{\lambda}^+ = \hat{\sigma}^2(1+\sqrt{\gamma})^2, \quad \gamma = C_r/N, \tag{1}$$

and the signal dimensionality and projection were

$$k_r(t) = \max\big(1, |\{i : \lambda_i > \hat{\lambda}^+\}|\big), \qquad \mathbf{S}_r(t) = \mathbf{X}_r(t)\mathbf{V}_r^{(k)} \in \mathbb{R}^{N\times k_r(t)}, \tag{2}$$

where $\mathbf{V}_r^{(k)}$ contains the top $k_r(t)$ eigenvectors of $\hat{\mathbf{\Sigma}}_r$; a floor of one component is retained to avoid degenerate representations at bins with no detectable signal. In our regime ($\gamma \in [0.002, 0.10]$; V1: C=448, V4/IT: C=256, $N \geq$22,000), signal components occupy at most 10–30% of the eigenspectrum, so the median falls within the MP noise bulk [21]; any upward bias in $\hat{\sigma}^2$ raises $\hat{\lambda}^+$ and under-retains signal, which would tend to reduce, rather than inflate, retained signal (Section 2.4).

Joint representations for synergy estimation were obtained by concatenating per-ROI projections $\{\mathbf{S}_r(t)\}$ column-wise and applying a second-level MP-PCA pass to the combined matrix, using the same threshold procedure with updated aspect ratio $\gamma_{\mathrm{joint}} = \sum_r k_r(t)/N$. This step removes residual noise components that survive single-area filtering and balances contributions from areas with differing electrode counts and intrinsic dimensionalities, so that the resulting joint projection $\mathbf{Z}(t)$ is intended to capture inter-area shared structure. A dedicated bias-correction term was estimated separately for $\mathbf{Z}(t)$ (Section 2.4).

2.4 Mutual Information Estimation under a Gaussian Model

Let $\mathbf{S} \in \mathbb{R}^{N \times d_S}$ and $\mathbf{Y} \in \mathbb{R}^{N \times d_Y}$ be zero-mean neural and target matrices. Under a jointly Gaussian assumption, the mutual information admits the closed-form log-determinant expression

$$\hat{I}(\mathbf{S}; \mathbf{Y}) \;=\; \frac{1}{2}\Big(\log|\hat{\mathbf{\Sigma}}_S| + \log|\hat{\mathbf{\Sigma}}_Y| - \log|\hat{\mathbf{\Sigma}}_{SY}|\Big), \tag{3}$$

where $\hat{\mathbf{\Sigma}}_{SY}$ is the joint sample covariance of $[\mathbf{S}, \mathbf{Y}]$ and log-determinants are computed from signed LU factorisation; the estimate is clipped to zero if any determinant sign is non-positive.

The Gaussian plug-in estimator was chosen on theoretical and practical grounds. Under a jointly Gaussian model, mutual information depends only on second-order statistics and thus quantifies the information accessible to optimal linear decoders—an operational proxy for inter-area communication under biologically plausible readouts [30–33]. Accordingly, it measures linear-accessible (second-order) information rather than total information; higher-order dependencies not reflected in covariance are not captured by design. This aligns with MP-PCA, which emphasizes shared variance across electrodes. Practically, the Gaussian framework remains well-conditioned and tractable in the 50–160D, large-scale regime considered (57 time bins $\times$ 5 CNN layers $\times$ 500 shuffles), where k-NN and neural estimators are statistically unstable and computationally prohibitive [22]. As a sensitivity check, we repeated the analysis with a Gaussian Copula MI (GCMI) estimator on a 3,000-sample subset (Supplementary Figs. S2–S4). Absolute magnitudes differed, and temporal profiles correlated with the plug-in estimates at $r \approx 0.75$ for synergy and ≈ 0.9 for redundancy and unique information, so synergy is the least stable atom across estimators.

Shuffle-based bias correction. Finite-sample bias was corrected by subtracting a shuffle-based empirical mean: for each time bin, $\mathbf{Y}$ was row-permuted $B = 500$ times on a subsample of $N_{\text{sub}} = 5{,}000$ trials, and the bias estimate $\hat{b}$ was the mean MI across permutations. Separate bias terms were estimated for each individual-ROI MI term and for each joint representation; for pairwise analyses, the per-area marginal bias terms were taken from the three-source estimation (the marginal MI of a single area is independent of the other sources present), while a dedicated joint bias was estimated for each pair's joint projection. All bias-corrected estimates satisfy $\hat{I}^{\text{corr}} = \max(0,\, \hat{I} - \hat{b})$. All PID time series were additionally baseline-corrected by subtracting the pre-stimulus mean (bins centred at < 100 ms); the same column-wise subtraction was applied independently to each shuffle replicate before any threshold computation, ensuring observed and null values are compared on a consistent change-from-baseline scale.

2.5 Partial Information Decomposition

Given neural source representations $\{\mathbf{S}_r\}_{r=1}^{n}$ and target $\mathbf{Y}$, Partial Information Decomposition [24] decomposes the total joint information into non-negative atoms of redundancy, unique information, and synergy.

Redundancy and unique information. We use minimum-mutual-information (MMI) redundancy [23]. For jointly Gaussian systems with a scalar target, axiom-satisfying decompositions reduce to MMI; our targets are 50-dimensional, so we adopt it as a definitional choice. It differs from the Williams–Beer $I_{\min}$. Redundancy is defined as the minimum information any single source carries about the target:

$$R \;=\; \min_{r}\; \hat{I}^{\text{corr}}(\mathbf{S}_r;\, \mathbf{Y}). \tag{4}$$

The unique information of source r is the excess above this redundant floor:

$$U_r \;=\; \max\big(0,\; \hat{I}^{\text{corr}}(\mathbf{S}_r;\, \mathbf{Y}) - R\big). \tag{5}$$

Synergistic information. Synergy is the information present in the joint representation that cannot be accounted for by the redundant and unique components:

$$\text{Syn} \;=\; \max\Big(0,\; \hat{I}^{\text{corr}}\big(\mathbf{Z};\, \mathbf{Y}\big) - R - \textstyle\sum_r U_r\Big), \tag{6}$$

where $\mathbf{Z}$ is the second-level MP-PCA projection of the concatenated source representations (Section 2.3). The $\max(0, \cdot)$ operations enforce non-negativity of all atoms, following the lattice construction of Williams and Beer [24] applied to the Gaussian framework of Barrett [23]. Because the

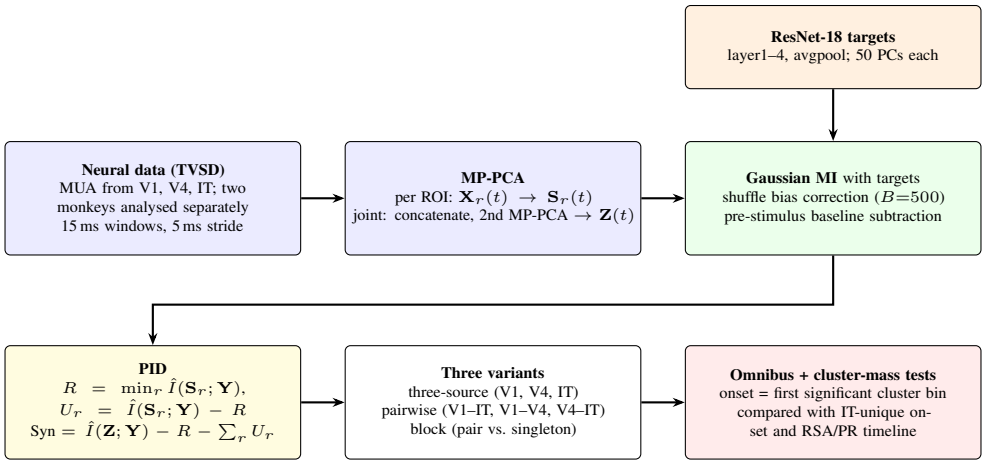

Figure 1: **Analysis pipeline.** MUA from V1, V4 and IT is windowed and reduced by Marchenko–Pastur PCA per area, then jointly. Gaussian mutual information with ResNet-18 targets is bias-corrected and baseline-subtracted, then decomposed into redundancy, unique and synergistic atoms in three-source, pairwise and block configurations. Onsets are tested with omnibus and cluster-mass permutation tests and compared with the RSA/PR feedforward reference timeline. Robustness: five window/stride configurations.

joint bias $\hat{b}_{\mathbf{Z}}$ is estimated separately from the marginal biases $\{\hat{b}_r\}$, the decomposition avoids the systematic over-correction of the synergy atom that arises from applying a single pooled bias term.

Our conclusions rest on relative ordering and onset timing rather than absolute magnitudes. These are not invariant to every estimator choice, because each atom combines several MI terms and is clipped at zero, so we assess sensitivity directly (GCMI check; window/stride sweep).

Three-source hybrid PID. For the three-area (V1, V4, IT) decomposition, sources were $\{\mathbf{S}_{\mathrm{V1}}, \mathbf{S}_{\mathrm{V4}}, \mathbf{S}_{\mathrm{IT}}\}$. Redundancy was $R = \min_r \hat{I}^{\mathrm{corr}}(\mathbf{S}_r; \mathbf{Y})$, unique information was computed per area, and synergy was computed against the joint second-level projection $\mathbf{Z}_{\mathrm{all}}$ of all three sources.

Pairwise PID. For each of the three area pairs $(r_A, r_B) \in \{(\mathrm{V1}, \mathrm{IT}),\ (\mathrm{V1}, \mathrm{V4}),\ (\mathrm{V4}, \mathrm{IT})\}$, an independent two-source PID was computed using $\mathbf{S}_{r_A}$, $\mathbf{S}_{r_B}$, and their dedicated second-level joint projection $\mathbf{Z}_{r_A r_B}$.

Block PID. To directly test the temporal ordering of feedforward integration—specifically whether the V1+V4 block, representing the canonical feedforward input to IT, synergises with IT before or after other area combinations—pairs of areas were treated as a single composite source. For each of the three block configurations—$\{(\mathrm{V1{+}IT\ vs.\ V4}),\ (\mathrm{V1{+}V4\ vs.\ IT}),\ (\mathrm{V4{+}IT\ vs.\ V1})\}$—the pair block was represented by the second-level MP-PCA projection of its two constituent areas' individual signal projections, following the same procedure as the three-source joint representation described above. A two-source PID was then computed between this block projection and the singleton area's signal projection.

2.6 Statistical, Representational, and Robustness Analyses

For Total MI, Synergy, and pairwise/block components, two permutation tests were applied: an omnibus test comparing mean post-stimulus (≥ 100 ms) values against $B = 500$ shuffle means, and a cluster-mass test using a threshold of $\mu_{\mathrm{null}}(t) + 2\sigma_{\mathrm{null}}(t)$ and a maximum-cluster-mass null distribution ($\alpha = 0.05$). Cluster onset, defined by the first significant cluster bin, provides only an approximate lower bound on effect onset [28]. Onsets are reported only for series in which both the omnibus and cluster-mass tests were significant. RSA [27] compared each area's MP-PCA projections with CNN targets via Spearman correlation of Euclidean-distance RDMs on 500-trial subsamples (5 repetitions), with 200 label permutations ($\alpha = 0.05$); the earliest significant bin defined representational onset and the reference for synergy-onset comparisons. The participation ratio of



Figure 2: **Sequential feedforward sweep in both animals.** *Top row:* RSA alignment (Spearman r) between each area's MP-PCA projections and AvgPool CNN features. Onset follows V1→V4→IT in both animals (5 ms apart in Monkey N). *Bottom row:* Participation ratio (PR) quench follows the same order in Monkey F; in Monkey N, V1 and V4 coincide.

each area was $\mathrm{PR} = (\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$ over its covariance eigenspectrum. Robustness was assessed across $(W, s) \in \{(10, 2), (15, 4), (20, 5), (30, 5), (50, 10)\}$ ms using 5,000 trials per configuration; consistent synergy onset and peak magnitude confirmed stability of synergy onset and peak magnitude (Supplementary Fig. S1).

3 Results

We used simultaneous MUA recordings from V1, V4, and IT in two macaque monkeys (Monkey F: 512 V1, 320 IT, 192 V4 electrodes; Monkey N: 448 V1, 256 IT, 256 V4 electrodes), each viewing 25,248 natural-image trials from the THINGS dataset collected by [15]. We applied Partial Information Decomposition (PID) with per-area Marchenko–Pastur PCA (MP-PCA) noise rejection in 15 ms windows stepped by 5 ms, targeting four distinct ResNet-18 representations (Layer 4 and AvgPool coincide) spanning low-level spatial features (Layer 1, 64-d) to high-level object identity (AvgPool, 512-d). Our headline claims are stated at the AvgPool target (512-d object identity); the Layer 1–3 sweep is reported as a secondary analysis of how the results vary with target level, and wherever a result is quoted at another layer we give the layer, area pair and time point alongside it. Statistical significance was assessed using omnibus and cluster-mass permutation tests with 500 label shuffles. Synergy onset and peak magnitude were stable across five window–stride configurations (Supplementary Fig. S1) .

Overview of findings. Three results emerge, each replicating across both animals. First, the ventral stream shows the expected serial sweep: representational content appears in the order V1→V4→IT. Second, alongside this sweep, the joint V1–V4–IT population carries synergistic information that begins before IT carries any unique information of its own. Third, within this joint code, the non-adjacent V1–IT pair couples at least as strongly as, at mid-to-high-level targets, the adjacent pairs, while the combined V1+V4 block joins the synergy with IT last. Taken together, the pairwise and block results indicate that serial and joint codes coexist in the same recordings. Per-layer values are given in Supplementary Tables S1–S3.

3.1 Sequential feedforward sweep

As shown in Figure 2, in Monkey F RSA onset and PR quench follow V1→V4→IT (RSA onsets 122.5, 137.5, 152.5 ms). In Monkey N, RSA onsets are ordered the same way but only 5 ms apart (127.5, 132.5, 137.5 ms), and V1 and V4 PR quench coincide (142.5 ms; Supplementary Table S1). These orderings provide the reference timeline against which synergy onset is compared below.

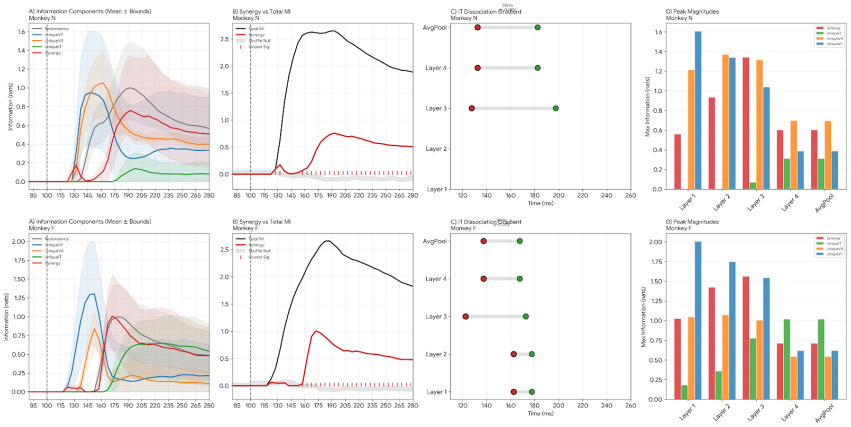

Figure 3: **Synergy precedes IT-unique information.** (A) PID components (Layer 4); synergy (red) rises before IT-unique (green). (B) Synergy (red dashed) and total MI (black) at AvgPool with shuffle null (grey). (C) Synergy onset (red) and IT-unique onset (green) per layer and animal; gaps are 15–70 ms, with the same direction in every condition where IT-unique information is detected. (D) Peak magnitudes per layer.

3.2 Inter-area synergy begins before IT carries unique object information

The serial feedforward account predicts that individual area representations should precede inter-area coupling. We find the opposite for IT in every condition where IT-unique information is detected.

Synergy — information accessible only from the joint population and absent from any single area — is statistically significant at all layers in both animals (cluster-mass p=0.002, omnibus p=0.002). It emerges substantially before IT unique information at every measurable layer. At AvgPool specifically: synergy onset (first significant cluster bin) is 137.5 ms (Monkey F) and 132.5 ms (Monkey N), while IT unique information onset is 167.5 ms (Monkey F) and 182.5 ms (Monkey N) — gaps of 30 ms and 50 ms respectively. Across all measurable layers the gap ranges from 15–70 ms: smallest at Layers 1–2 in Monkey F (15 ms), largest at Layer 3 in Monkey N (70 ms), shrinking from Layer 3 to AvgPool in both animals (Figure 3C); full per-layer onset timings are given in Supplementary Table S3. IT unique information is absent in Monkey N for Layers 1–2, possibly reflecting Monkey N's smaller IT array.

As shown in Figure 3B and D, Synergy magnitude peaks at 1.562 nats in Monkey F and 1.343 nats in Monkey N at Layer 3, and at 0.712 nats and 0.603 nats at AvgPool. At AvgPool, synergy rises sharply to its peak within the first 40 ms post-onset and then decays to a sustained plateau — remaining above 0.15 nats from 162.5 ms to the end of the window in Monkey F and above 0.44 nats from 172.5 ms onwards in Monkey N — suggesting that synergy is not only transient. V1–IT pairwise synergy sustains at 0.244 nats for over 120 ms *after* V1's unique information reaches zero (at 167.5 ms) while IT-unique information continues to rise (peaking at 1.018 nats at 237.5 ms, Monkey F).

3.3 Near-parallel, non-serial onset of inter-areal synergistic coupling

Anatomical adjacency predicts that directly adjacent pairs, specifically V1–V4 and V4–IT, and then recurrent feedback, plausibly V1–IT in that order, should proceed sequentially in terms of pairwise synergistic coupling. However, we observe that the pairwise synergies V1–IT and V1–V4 begin simultaneously and also start their ascent near-simultaneously. V1–IT and V1–V4 share the same feedforward onset timing—0 ms apart in Monkey N (all pairs at 132.5 ms) and 5 ms apart in Monkey F (V1–V4 at 122.5 ms, V1–IT at 127.5 ms)—and rise in parallel throughout the steep post-onset phase. This suggests parallel collaborative coding of the image representation. We expected adjacent pairs to carry more synergy than the non-adjacent V1–IT pair; this was not the case at the object-identity target in either animal. At AvgPool, V1–IT had the largest peak synergy (Monkey F: V1–IT 0.587 $>$ V4–IT 0.539 $>$ V1–V4 0.511 nats; Monkey N: V1–IT 0.582 $>$ V4–IT 0.424 $>$ V1–V4 0.393 nats). V1–IT was also the largest pair at Layer 3 and AvgPool in both animals, and at all four distinct layers in Monkey F (Figure 4). At the two lowest layers in Monkey N, V1–V4 was larger (0.774/0.900 nats).

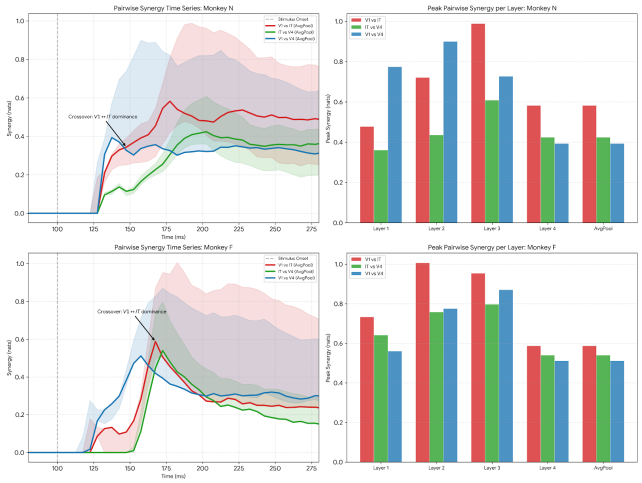

Figure 4: **Pairwise synergy emerges near-concurrently** *Left:* Pairwise synergy time series (Layer 4) for each monkey. V1↔IT (red) and V1↔V4 (blue) rise in parallel before V1↔IT takes the lead at 167.5 ms (Monkey F, annotated arrow). *Right:* Peak pairwise synergy across all five CNN layers. V1↔IT is strongest at all layers in Monkey F and from Layer 3 onward in Monkey N.

The three pairs emerge near-concurrently: the onset spread at AvgPool is 0 ms in Monkey N and 30 ms in Monkey F (122.5–152.5 ms). With a 15 ms window, 5 ms stride and cluster onsets that give only an approximate lower bound [28], we read this as near-concurrent rather than excluding a short serial lag.

3.4 The V1+V4 block couples with IT last

Under a feedforward account, synergy between the V1+V4 block and IT can only appear once IT carries information, so a late block–IT coupling is expected and is not discriminating on its own. We read it alongside Section 3.3, where V1–IT coupling emerges at response onset, well before block–IT coupling. V1–V4 coupling begins at or slightly before V1–IT coupling (Section 3.3; Supplementary Table S5), so we do not infer the order in which V4 joins, or whether V4 mediates V1–IT coupling (a direct V2/V3 route to posterior IT exists alongside a more important V4 route [36]). At AvgPool in Monkey F the block onset (167.5 ms) coincides with the IT-unique onset.

V1+V4 vs. IT block synergy emerged last in every reported condition . In Monkey F, the onset ordering at Layers 1–3 is V1+IT (142.5 ms) → V4+IT (147.5 ms) → V1+V4 (162.5 ms), and at Layer 4 and AvgPool it is V1+IT (152.5 ms) = V4+IT (152.5 ms) → V1+V4 (167.5 ms). In Monkey N, the ordering at Layers 1–3 is V4+IT (132.5 ms) → V1+IT (137.5 ms) → V1+V4 (152.5 ms at Layers 1–2; 147.5 ms at Layer 3), and at Layer 4 and AvgPool it is V1+IT (142.5 ms) → V4+IT (147.5 ms) → V1+V4 (167.5 ms). V1+V4 vs. IT is latest in all ten reported conditions (eight distinct, since Layer 4 and AvgPool coincide), by 10–20 ms (2–4 strides; the window is 15 ms; Figure 5).

3.5 Control for dimensionality

MP-PCA retains a different number of components per area and window, and joint sources have more dimensions than single areas, which could favour the joint representation. We repeated the full PID with each source's dimensionality fixed at k=60 (Monkey N; ResNet-18 layer2 and avgpool). Three-source synergy was again nonzero from 137.5–142.5 ms, 30–35 ms before IT-unique information (172.5 ms). The pairwise pattern was again target-dependent: at avgpool, V1–IT had the largest peak synergy (0.432 vs. 0.391 for V4–IT and 0.281 for V1–V4), whereas at layer2 V1–V4 was largest (0.618 vs. 0.565 for V1–IT and 0.440 for V4–IT; Supplementary Table S9). This control covers one animal and two targets, and it does not test the persistence of V1–IT synergy after V1-unique information reaches zero.

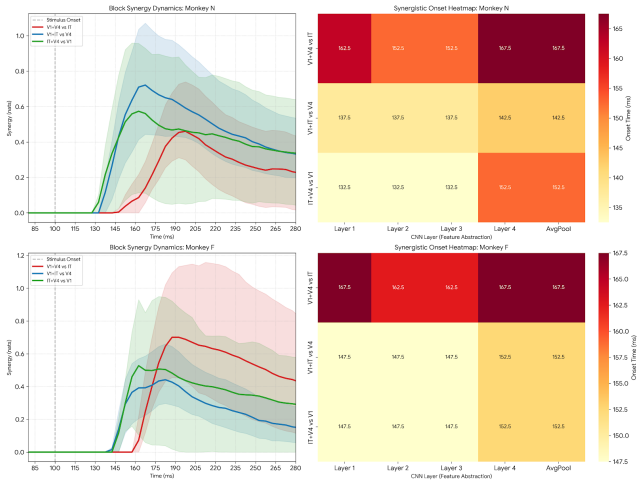

Figure 5: **The V1+V4 block couples with IT last in all ten reported conditions.** *Left:* Block synergy time series (Layer 4). V1+V4 vs. IT (red) rises last. *Right:* Onset heatmap across block configurations and CNN layers — V1+V4 vs. IT row is darkest (latest) in every cell.

Table 1: Cross-model results from the pairwise PID. Peak pairwise synergy in nats (largest pair in bold). Onsets in ms: first bin with nonzero V1–IT synergy, and first bin with nonzero IT-unique information (n.d. = not detected). DINO: Monkey N; CorNet-S: Monkey F.

		Onset (ms)		Peak synergy (nats)		
Model	Target	V1–IT syn.	IT-unique	V1–IT	V1–V4	V4–IT
DINO ViT-S/16	block 3	127.5	n.d.	0.604	**1.002**	0.407
DINO ViT-S/16	block 9	127.5	182.5	**0.789**	0.552	0.560
CorNet-S	V1	117.5	n.d.	**0.646**	0.524	0.599
CorNet-S	V4	122.5	172.5	**0.972**	0.882	0.783
CorNet-S	IT	127.5	167.5	**0.606**	0.495	0.519

3.6 Generalisation to other target models

For four of the five targets (DINO block 9 and all three CorNet-S areas), V1–IT has the largest peak pairwise synergy of the three area pairs (Table 1). The exception is the early DINO layer (block 3), where V1–V4 is largest, as for the low-level ResNet-18 targets in Monkey N. In the three targets in which IT-unique information was detected (DINO block 9, CorNet-S V4, CorNet-S IT), V1–IT synergy became nonzero 40–55 ms before IT-unique information. In the other two targets no IT-unique information was detected, so no comparison is possible. Each model was run in one animal, margins are small in places (CorNet-S V1: 0.646 vs. 0.599 nats), no confidence intervals were computed, and onsets are first nonzero bins, not significance-tested (Supplementary Table S8). CorNet-S areas as targets give V1–IT dominance, whereas CorNet-S as a source of V1/V4/IT (Section 3.7) does not; the former shows the ordering is not specific to ResNet-18, not that model-internal structure reproduces it.

3.7 Model-internal control: layers predicting layers

To ask whether synergy of this kind arises without neural data, we ran the same PID pipeline with Gaussianized model layers as sources and an external model as target (VGG-16 fc7 or DINOv2 CLS, to avoid circularity), using 6,000 THINGS images. Non-adjacent layer pairs in ResNet-18, VGG-16, MobileNetV2, SqueezeNet, and DINOv2 all showed nonzero synergy (0.31–0.45 nats for the CNNs with the VGG-16 fc7 target, 0.11–0.15 nats for three of them with the DINOv2 CLS target, 0.67–1.92 nats for DINOv2; Supplementary Table S6), so nonzero synergy is not specific to the brain. These runs used a single layer pair per model, so they do not test the ordering of pairs. CorNet-S, whose V1, V4, and IT areas are specified by the model's authors, allowed the three-pair comparison (Table 2): V4–IT had the largest synergy on both targets, V1–IT exceeded V1–V4, and

Table 2: CorNet-S internal PID (final timestep): synergy in nats with an external model as target.

Pair	VGG-16 fc7	DINOv2 CLS
V1–IT	0.757	0.268
V1–V4	0.432	0.061
V4–IT	**2.575**	**1.596**

V1–IT was the largest pair in 0 of 30 bootstrap resamples per target. Thus V1–IT dominance was not reproduced in CorNet-S. The second MP-PCA pass used for neural data was dropped for model sources, because two layers of one forward pass are deterministically related and the pass pruned the joint representation below either source (zero synergy for every model); this is expected to raise model synergy relative to the neural pipeline. Marginal information is also more unbalanced for V1–IT than V4–IT in CorNet-S (balance 0.18/0.19 vs. 0.53/0.66; Supplementary Table S7), so raw synergies are not directly comparable across pairs or with the neural data.

4 Discussion and conclusion

Our results are consistent with two concurrent regimes in the macaque ventral stream. First, a serial progression is visible: RSA onset and dimensionality quench follow V1→V4→IT, with small lags in Monkey N. Second, the joint V1–V4–IT population carries synergistic information that becomes measurable before IT-unique information, and at mid-to-high-level targets the non-adjacent V1–IT pair carries at least as much pairwise synergy as the adjacent pairs. The V1+V4 block couples with IT last, which a feedforward account also predicts; the pairwise and block results therefore do not discriminate between mechanisms, and they do not identify whether V4 mediates V1–IT coupling [36]. Because the ventral stream is known to be recurrent [37] and feedforward networks also show nonzero synergy under our measure, synergy here is a statistical signature of joint structure, not evidence of a specific circuit. Models of the ventral stream could be evaluated on multi-area information structure in addition to per-area encoding. We did not train a model to reproduce this coupling: optimising the measured quantity would satisfy it by construction, and we lack an account of which architectural property would yield it. Our contribution is diagnostic, and we cannot claim that such coupling would improve task performance.

Limitations. We use a Gaussian plug-in MI estimator on MP-PCA projections, which captures second-order structure; reported synergy is linearly accessible joint structure, not total information. Our PID uses MMI redundancy, a definitional choice for 50-dimensional targets that may assign more mass to synergy than alternatives such as I_{dep} or BROJA, so absolute synergy should be read cautiously. Our findings concern relative ordering and timing, which are not invariant to estimator choice: in the GCMI check, synergy profiles correlated with the plug-in estimates at only ≈ 0.75, and we did not test other redundancy measures. Targets were mainly a pretrained ResNet-18; DINO ViT-S/16 and CorNet-S were tested in one animal each without confidence intervals, and the model-internal control used one recurrent architecture, a 6,000-image subset, a pipeline without the second MP-PCA pass, and unbalanced marginals across CorNet-S pairs.RSA correlations are low in absolute terms; spatiotemporal RSA has been applied to human MEG–fMRI data [38], though with a different measure, so magnitudes are not directly comparable. The split-half noise ceiling (0.24–0.44) was computed on 100 repeated images with 15-trial averages and does not bound the single-trial values reported here. The persistence of V1–IT synergy after V1-unique information reaches zero was not tested under fixed dimensionality. Finally, passive viewing does not engage top-down attention, so whether this structure changes under task demands is open.

5 Data Availability

The neural data analysed here are the publicly available TVSD recordings of Papale et al. [15] (Neuron 113(4), 2025; DOI 10.1016/j.neuron.2024.12.003; data at `https://doi.gin.g-node.org/10.12751/g-node.hc7zlv/`). We did not collect these data. Instructions for running our code on the dataset are in the accompanying `README.md`.

References

[1] DiCarlo, J.J., Zoccolan, D., & Rust, N.C. (2012) How does the brain solve visual object recognition? *Neuron* **73**(3):415-434. doi:10.1016/j.neuron.2012.01.010.

[2] Yamins, D.L. & DiCarlo, J.J. (2016) Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience* **19**(3):356-365. doi:10.1038/nn.4244.

[3] Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., & DiCarlo, J.J. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111**(23):8619–8624. doi:10.1073/pnas.1403112111.

[4] Khaligh-Razavi, S.M. & Kriegeskorte, N. (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Computational Biology* **10**(11):1–29. doi:10.1371/journal.pcbi.1003915.

[5] Güçlü, U. & van Gerven, M.A.J. (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience* **35**(27):10005–10014. doi:10.1523/JNEUROSCI.5023-14.2015.

[6] Nayebi, A., Sagastuy-Brena, J., Bear, D.M., Kar, K., Kubilius, J., Ganguli, S., Sussillo, D., DiCarlo, J.J., & Yamins, D.L. (2022) Recurrent connections in the primate ventral visual stream mediate a trade-off between task performance and network size during core object recognition. *Neural Computation* **34**(8):1652–1675. doi:10.1162/neco_a_01506.

[7] Nayebi, A., Bear, D., Kubilius, J., Kar, K., Ganguli, S., Sussillo, D., DiCarlo, J.J., & Yamins, D.L. (2018) Task-driven convolutional recurrent models of the visual system. *Advances in Neural Information Processing Systems* **31**.

[8] Kar, K. & DiCarlo, J.J. (2021) Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *Neuron* **109**(1):164–176. doi:10.1016/j.neuron.2020.09.035.

[9] Williams, P.L. & Beer, R.D. (2010) Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*.

[10] Kar, K., Kubilius, J., Schmidt, K., Issa, E.B., & DiCarlo, J.J. (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience* **22**(6):974–983. doi:10.1038/s41593-019-0392-5.

[11] Ercsey-Ravasz, M., Markov, N.T., Lamy, C., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013) A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* **80**(1):184–197. doi:10.1016/j.neuron.2013.07.036.

[12] Luppi, A.I., Mediano, P.A.M., Rosas, F.E., Holland, N., Fryer, T.D., O'Brien, J.T., Rowe, J.B., Menon, D.K., Bor, D., & Stamatakis, E.A. (2022) A synergistic core for human brain evolution and cognition. *Nature Neuroscience* **25**(6):771–782. doi:10.1038/s41593-022-01070-0.

[13] Gelens, F., Äijälä, J., Roberts, L., et al. (2024) Distributed representations of prediction error signals across the cortical hierarchy are synergistic. *Nature Communications* **15**:3941. doi:10.1038/s41467-024-48329-7.

[14] Roberts, L., Äijälä, J., Burger, F., et al. (2026) Broadband synergy versus oscillatory redundancy in the visual cortex. *Nature Communications*. doi:10.1038/s41467-026-72444-2.

[15] Papale, P., Wang, F., Self, M.W., & Roelfsema, P.R. (2025) An extensive dataset of spiking activity to reveal the syntax of the ventral stream. *Neuron* **113**(4):539–553.e5. doi:10.1016/j.neuron.2024.12.003.

[16] Hung, C.P., et al. (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* **310**(5749):863–866. doi:10.1126/science.1117593.

[17] Panzeri, S., Moroni, M., Safaai, H., & Harvey, C.D. (2022) The structures and functions of correlations in neural population codes. *Nature Reviews Neuroscience* **23**(9):551–567. doi:10.1038/s41583-022-00606-4.

[18] Hebart, M.N., Dickter, A.H., Kidder, A., Kwok, W.Y., Corriveau, A., Van Wicklin, C., & Baker, C.I. (2019) THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE* **14**(10):1–24. doi:10.1371/journal.pone.0223792.

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016) Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*. doi:10.48550/arXiv.1512.03385.

[20] Marčenko, V.A. & Pastur, L.A. (1967) Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* **1**(4):457–483. doi:10.1070/SM1967v001n04ABEH001994.

[21] Gavish, M. & Donoho, D.L. (2014) The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory* **60**(8):5040–5053. doi:10.1109/TIT.2014.2323359.

[22] Ince, R.A.A., Giordano, B.L., Kayser, C., Rousselet, G.A., Gross, J., & Schyns, P.G. (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Human Brain Mapping* **38**(3):1541–1573. doi:10.1002/hbm.23471.

[23] Barrett, A.B. (2015) Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E* **91**(5):052802. doi:10.1103/PhysRevE.91.052802.

[24] Williams, P.L. & Beer, R.D. (2010) Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*.

[25] Gelens, F., Äijälä, J., Roberts, L., et al. (2024) Distributed representations of prediction error signals across the cortical hierarchy are synergistic. *Nature Communications* **15**:3941. doi:10.1038/s41467-024-48329-7.

[26] Roberts, L., Äijälä, J., Burger, F., et al. (2026) Broadband synergy versus oscillatory redundancy in the visual cortex. *Nature Communications*. doi:10.1038/s41467-026-72444-2.

[27] Kriegeskorte, N., Mur, M., & Bandettini, P. (2008) Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2**:4. doi:10.3389/neuro.06.004.2008.

[28] Sassenhagen, J. & Draschkow, D. (2019) Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* **56**(6):e13335. doi:10.1111/psyp.13335.

[29] Markov, N.T., Ercsey-Ravasz, M.M., Ribeiro Gomes, A.R., Lamy, C., Magrou, L., Vezoli, J., ... & Kennedy, H. (2014) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex* **24**(1):17–36. doi:10.1093/cercor/bhs270.

[30] Cover, T.M. & Thomas, J.A. (2006) *Elements of Information Theory*, 2nd ed. John Wiley & Sons. doi:10.1002/047174882X.

[31] Shamir, M. & Sompolinsky, H. (2006) Implications of neuronal diversity on population coding. *Neural Computation* **18**(8):1951–1986. doi:10.1162/neco.2006.18.8.1951.

[32] Pouget, A., Dayan, P., & Zemel, R. (2000) Information processing with population codes. *Nature Reviews Neuroscience* **1**(2):125–132. doi:10.1038/35039062.

[33] Cunningham, J.P. & Yu, B.M. (2014) Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience* **17**:1500–1509. doi:10.1038/nn.3776.

[34] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021) Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

[35] Kubilius, J., Schrimpf, M., Kar, K., et al. (2019) Brain-like object recognition with high-performing shallow recurrent ANNs. *Advances in Neural Information Processing Systems* **32**.

[36] Ponce, C.R., Lomber, S.G., & Livingstone, M.S. (2017) Posterior inferotemporal cortex cells use multiple input pathways for shape encoding. *Journal of Neuroscience* **37**(19):5019–5034. doi:10.1523/JNEUROSCI.2674-16.2017.

[37] Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., & Mishkin, M. (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences* **17**(1):26–49. doi:10.1016/j.tics.2012.10.011.

[38] Wu, Y.-H., Podvalny, E., & He, B.J. (2023) Spatiotemporal neural dynamics of object recognition under uncertainty in humans. *eLife* **12**:e84797. doi:10.7554/eLife.84797.

A Gaussian Copula Mutual Information PID

To decompose information carried by visual cortical areas V1, IT, and V4 about stimulus identity, we applied Partial Information Decomposition using a Gaussian Copula Mutual Information (GCMI) estimator.

Copula estimator

Raw neural responses in each ROI were first reduced via MP-PCA (as in the main paper).The main analyses apply the Gaussian plug-in estimator directly to the MP-PCA components. As a sensitivity check that does not rely on marginal Gaussianity, we applied a rank-based normalization to each marginal prior to MI estimation. Specifically, for a variable $\mathbf{X} \in \mathbb{R}^{n \times d}$, each column was transformed via the probit function:

$$\tilde{X}_{ij} = \Phi^{-1}\left(\frac{r_{ij} - 0.5}{n}\right), \tag{S1}$$

where r_{ij} is the rank of X_{ij} within column j and Φ^{-1} is the standard normal quantile function. This marginal Gaussianization preserves the rank-order dependence structure while rendering the Gaussian MI formula applicable. MI was then estimated from covariances of the transformed data:

$$\widehat{I}(X;Y) = \frac{1}{2}\left(\log|\Sigma_X| + \log|\Sigma_Y| - \log|\Sigma_{XY}|\right), \tag{S2}$$

where Σ_X, Σ_Y, and Σ_{XY} are the covariance matrices of $\tilde{\mathbf{X}}$, $\mathbf{Y}$, and their concatenation, respectively. Both the neural component $\mathbf{X}$ and the CNN feature targets $\mathbf{Y}$ (PCA-50 of ResNet-18 avgpool activations) were Gaussianized.

PID decomposition

Total MI from the three ROIs jointly was decomposed into redundancy, unique, and synergistic atoms following the minimum-MI redundancy axiom:

$$\mathrm{Red} = \min_{s \in \{V1,\, IT,\, V4\}} \widehat{I}(S_s;\, Y), \tag{S3}$$

$$\mathrm{Uniq}_s = \max\Big(0,\ \widehat{I}(S_s;\, Y) - \mathrm{Red}\Big), \tag{S4}$$

$$\mathrm{Syn} = \max\Big(0,\ \widehat{I}(Z_{\mathrm{joint}};\, Y) - \mathrm{Red} - \textstyle\sum_s \mathrm{Uniq}_s\Big). \tag{S5}$$

B Computational Resources

All experiments and analyses were conducted using Google Colaboratory (Colab). Computations were performed utilizing an NVIDIA A100 Tensor Core GPU with the High-RAM runtime configuration. Model feature extraction for the model-internal analyses used an NVIDIA L4 GPU. This hardware environment provided the necessary memory capacity and compute capability to efficiently execute the models and process the neural population datasets. The source code required to reproduce all experimental results is included in our supplementary code submission.

Supplementary Table S1: Participation Ratio (PR) Quench Magnitudes and Temporal Latencies. Baseline PR is calculated as the mean dimensionality from 0–100 ms. Quench Time indicates the latency of the post-stimulus dimensionality minimum.

	Monkey N				**Monkey F**			
ROI	**Baseline** (PR)	**Min PR** (PR)	**Time** (ms)	**Quench %** (Δ%)	**Baseline** (PR)	**Min PR** (PR)	**Time** (ms)	**Quench %** (Δ%)
V1	34.47	11.20	142.5	67.51%	27.21	21.56	132.5	20.76%
V4	28.96	16.74	142.5	42.20%	27.03	25.88	147.5	4.26%
IT	24.57	18.16	192.5	26.10%	14.24	12.65	152.5	11.17%

Supplementary Table S2: Peak Informational Metrics across the Ventral Stream Hierarchy. $I(S;\mathbf{R})$ represents the total joint mutual information between the three cortical areas (V1, V4, IT) and the specific CNN layer representation. All values are reported in nats.

	Monkey N			Monkey F		
CNN Target Layer	**Peak** $I(S;\mathbf{R})$	**Peak** Syn	**Peak** Red	**Peak** $I(S;\mathbf{R})$	**Peak** Syn	**Peak** Red
ResNet18-Layer 1	3.25	0.56	0.97	2.85	1.02	1.09
ResNet18-Layer 2	3.34	0.94	1.24	3.17	1.42	1.38
ResNet18-Layer 3	3.42	1.34	1.49	3.43	1.56	1.43
ResNet18-Layer 4	1.94	0.60	0.66	1.97	0.71	0.55
ResNet18-AvgPool	1.94	0.60	0.66	1.97	0.71	0.55

Supplementary Table S3: Comparative Onset Latencies of Synergy and Unique IT Information. The Δ (ms) represents the temporal dissociation between circuit-level integrated activity and the settlement of independent IT representations.

	Monkey N			Monkey F		
CNN Target Layer	**Syn Onset** (ms)	**IT Onset** (ms)	Δ (ms)	**Syn Onset** (ms)	**IT Onset** (ms)	Δ (ms)
ResNet18-Layer 1	127.5	—	—	162.5	177.5	15.0
ResNet18-Layer 2	127.5	—	—	162.5	177.5	15.0
ResNet18-Layer 3	127.5	197.5	70.0	122.5	172.5	50.0
ResNet18-Layer 4	132.5	182.5	50.0	137.5	167.5	30.0
ResNet18-AvgPool	132.5	182.5	50.0	137.5	167.5	30.0

Supplementary Table S4: Comprehensive Peak PID Metrics across the ResNet-18 Hierarchy for Monkey N and Monkey F. All values are reported in nats.

	Monkey N						Monkey F					
Layer	**Red**	U_{V1}	U_{V4}	U_{IT}	**Syn**	I_{joint}	**Red**	U_{V1}	U_{V4}	U_{IT}	**Syn**	I_{joint}
Layer 1	0.97	1.61	1.22	0.00	0.56	3.25	1.09	2.01	1.05	0.18	1.02	2.85
Layer 2	1.24	1.34	1.37	0.00	0.94	3.34	1.38	1.75	1.07	0.36	1.42	3.17
Layer 3	1.49	1.04	1.32	0.07	1.34	3.42	1.43	1.54	1.00	0.78	1.56	3.43
Layer 4	0.66	0.39	0.70	0.31	0.60	1.94	0.55	0.62	0.54	1.02	0.71	1.97
AvgPool	0.66	0.39	0.70	0.31	0.60	1.94	0.55	0.62	0.54	1.02	0.71	1.97

Supplementary Table S5: Onset Latencies for Pairwise Synergy across the ResNet-18 Hierarchy. Values represent ms post-stimulus. Onsets are defined as in the main text (first significant bin; see Methods).

	Monkey N			Monkey F		
Layer	$V1 \leftrightarrow IT$	$V1 \leftrightarrow V4$	$V4 \leftrightarrow IT$	$V1 \leftrightarrow IT$	$V1 \leftrightarrow V4$	$V4 \leftrightarrow IT$
ResNet18-Layer 1	127.5	127.5	132.5	122.5	117.5	157.5
ResNet18-Layer 2	127.5	127.5	132.5	122.5	117.5	157.5
ResNet18-Layer 3	127.5	127.5	132.5	122.5	122.5	157.5
ResNet18-Layer 4	132.5	132.5	132.5	127.5	122.5	152.5
ResNet18-AvgPool	132.5	132.5	132.5	127.5	122.5	152.5

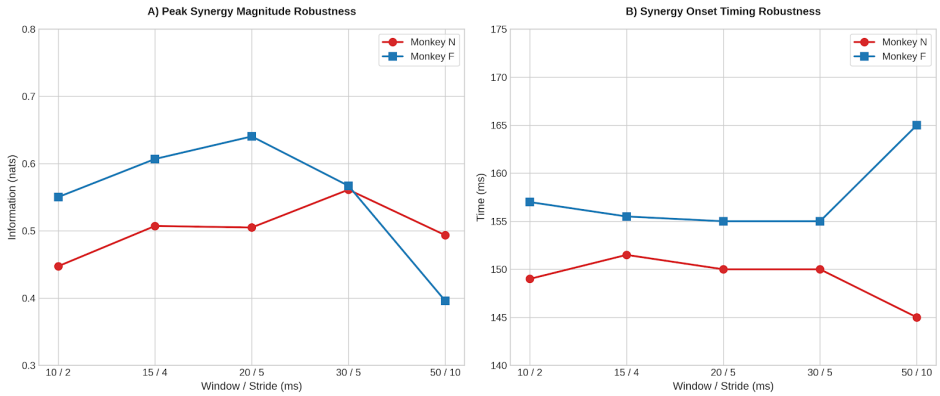

Figure S1: **(a) Stability of peak synergistic magnitude.** The maximum synergistic information (nats) remains robustly between 0.4 and 0.7 nats across both subjects (Monkey N, red circles; Monkey F, blue squares) for window lengths ranging from 10 ms to 50 ms. This invariance indicates that the synergistic component of the neural population code is a stable feature of the circuit's information topology, rather than an artifact of a specific integration timescale. **(b) Synergy onset across window settings.** In this analysis (5,000 trials per configuration; see Methods), the three-source synergy onset lies between about 145 ms and 165 ms across the parameter sweep in both animals. These absolute values differ from the onsets in Supplementary Table S3; only their stability across window settings is interpreted here.

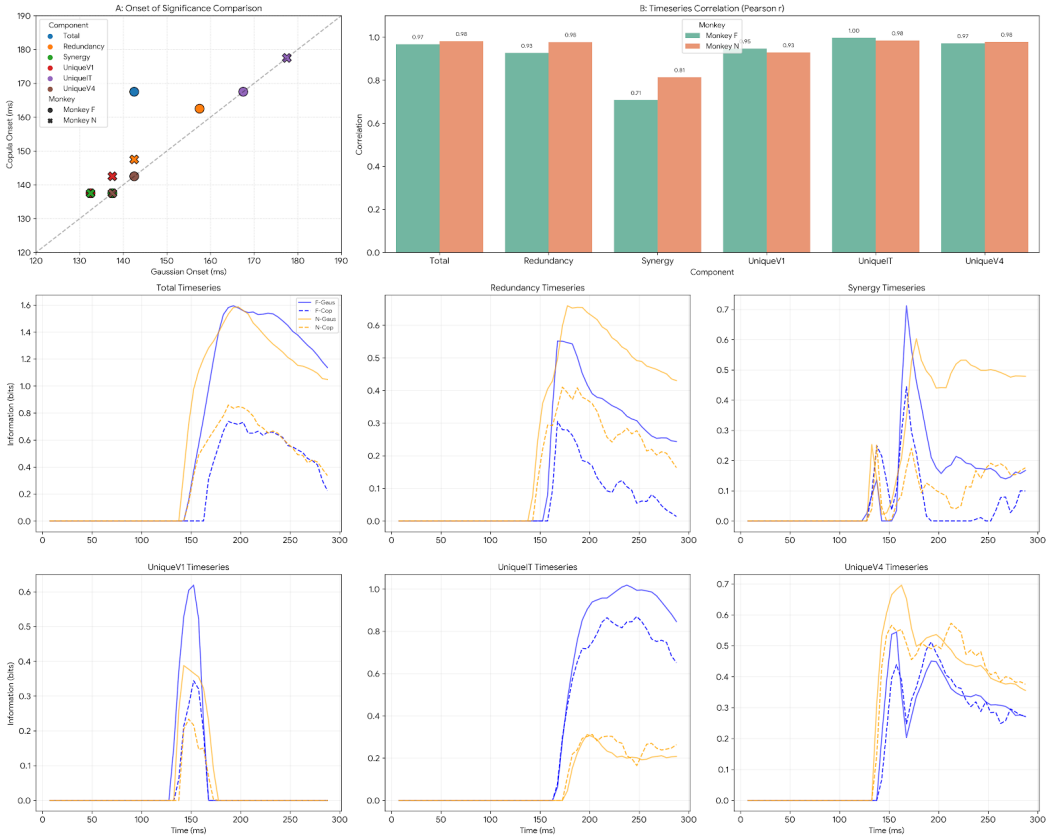

Figure S2: Validation of Gaussian Plug-in vs. Copula PID Estimators. Comprehensive comparison of the 3-source PID components across all monkeys. **(A) Onset of Significance Comparison.** Scatter plot illustrating high temporal consistency in the detection of information emergence across all PID atoms (Total MI, Redundancy, Synergy, and Unique information for V1, V4, and IT). **(B) Timeseries Correlation.** Pearson correlation coefficients (r) between the Gaussian and Copula timeseries for each PID component. Correlations are about 0.9 for redundancy and unique information and about 0.75 for synergy, so synergy is the least stable atom across estimators. **(Bottom Panels) Component Timeseries.** Representative temporal trajectories for each PID atom. Solid lines represent Gaussian plug-in estimates and dashed lines represent Copula estimates (Blue: Monkey F, Orange: Monkey N), highlighting the matched temporal envelopes across both monkeys and all estimators.

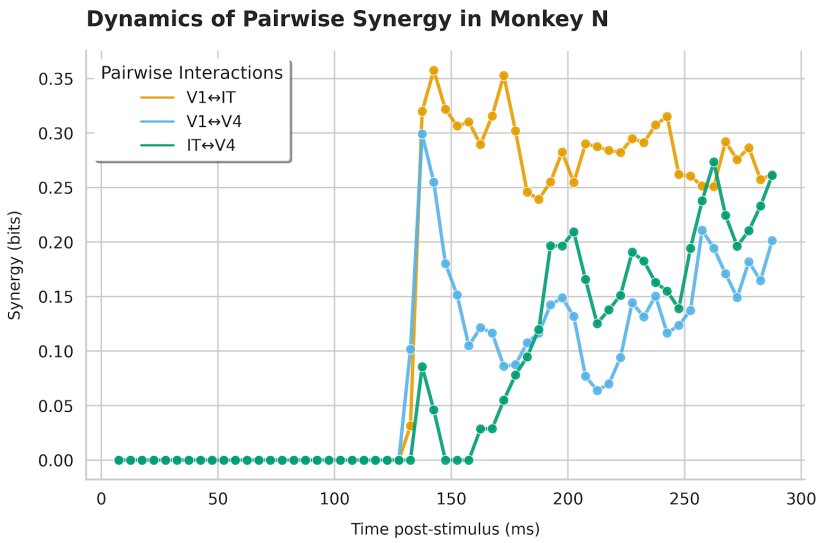

Figure S3: **Dynamics of Pairwise Synergy using GCMI**

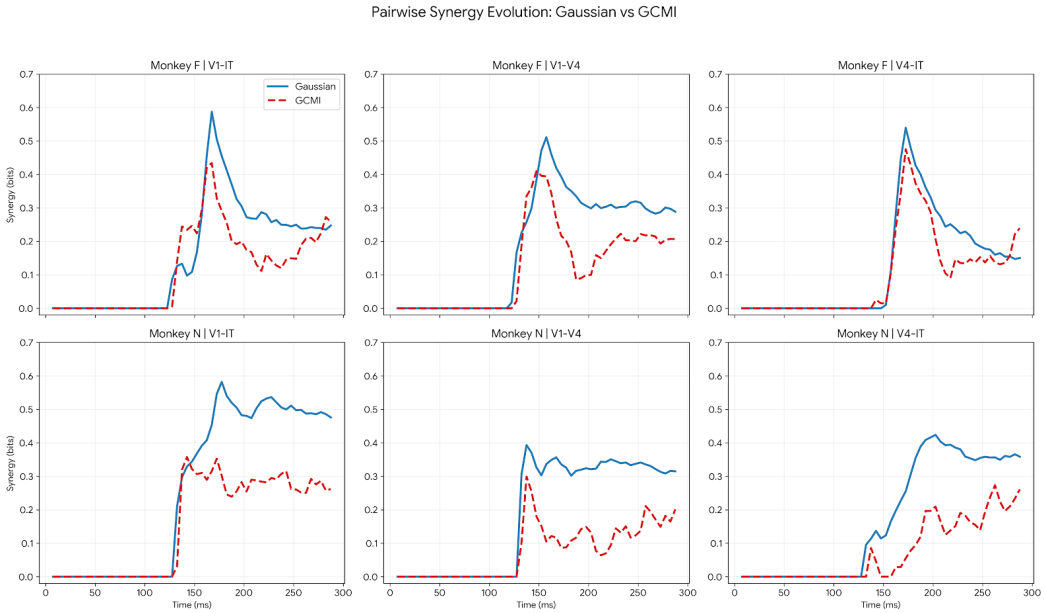

Figure S4: **Pairwise Synergy Evolution: Gaussian vs. GCMI.** Temporal dynamics of synergy calculated for cortical pairings (V1-IT, V1-V4, and V4-IT) across two subjects (Monkey F, top; Monkey N, bottom). Solid blue lines represent the Gaussian plug-in estimator (derived from spatially pooled Average Pool layer activations), while dashed red lines denote the GCMI (Copula-based) estimator. Both estimators show similar temporal profiles; GCMI magnitudes are lower, and profiles differ in detail.

Implementation. Features were extracted for 6,000 randomly sampled THINGS images, reduced by PCA to 50 components (31 for MobileNetV2 block 4), and passed through MP-PCA, shuffle-based bias correction, MMI redundancy, and synergy as in the neural pipeline, without a second MP-PCA pass on the concatenated sources. In CorNet-S, V4 has four recurrent timesteps and IT has two; "final" is the last timestep.

Table S6: non-adjacent layer pairs (synergy, nats).

Model	Layers	VGG-16 fc7 target	DINOv2 CLS target
ResNet-18	layer1–layer3	0.447	0.132
VGG-16	block2–block4	0.419†	–
MobileNetV2	block4–block10	0.308	0.152
SqueezeNet	fire3–fire9	0.420	0.111
DINOv2	block2–block6	0.673	–
DINOv2	block2–block8	1.259	–
DINOv2	block2–block10	1.269	–
DINOv2	block4–block8	1.638	–
DINOv2	block4–block10	1.593	–
DINOv2	block6–block10	1.916	–

† VGG-16 was run against ResNet-18 avgpool as target. DINOv2 was not run against the DINOv2 CLS target, to avoid circularity.

Table S7: CorNet-S per timestep (synergy, nats; marginal balance, where 1 is balanced).

	Synergy		Marginal balance	
Config	VGG-16 fc7	DINOv2 CLS	VGG-16 fc7	DINOv2 CLS
V1–IT (final)	0.757	0.268	0.180	0.192
V1–V4 (final)	0.432	0.061	0.337	0.290
V4–IT (final)	2.575	1.596	0.535	0.662
V1–V4, t1	0.263	0.010	0.541	0.476
V1–V4, t2	0.235	0.000	0.436	0.390
V1–V4, t3	0.423	0.068	0.389	0.343
V1–V4, t4	0.432	0.061	0.337	0.290
V1–IT, t1	0.470	0.076	0.241	0.222
V1–IT, t2	0.759	0.268	0.180	0.192

C Additional target models and dimensionality control

Table S8: all three pairs, per target. Onsets are the first time bin (ms) with nonzero pairwise synergy, not tested for significance. Peaks are in nats.

		Synergy onset (ms)			Peak synergy (nats)		
Model	Target	V1–IT	V1–V4	V4–IT	V1–IT	V1–V4	V4–IT
DINO ViT-S/16	block 3	127.5	127.5	132.5	0.604	1.002	0.407
DINO ViT-S/16	block 9	127.5	127.5	132.5	0.789	0.552	0.560
CorNet-S	V1	117.5	117.5	157.5	0.646	0.524	0.599
CorNet-S	V4	122.5	117.5	152.5	0.972	0.882	0.783
CorNet-S	IT	127.5	122.5	147.5	0.606	0.495	0.519

Table S9: dimensionality-matched control (k=60 per source; Monkey N). Three-source synergy was nonzero from 137.5–142.5 ms; IT-unique information became nonzero at 172.5 ms.

Target	Pair	Onset (ms)	Peak time (ms)	Peak value (nats)
Low-level (layer2)	V1–IT	132.5	197.5	0.565
Low-level (layer2)	V1–V4	127.5	207.5	0.618
Low-level (layer2)	V4–IT	132.5	197.5	0.440
High-level (avgpool)	V1–IT	137.5	182.5	0.432
High-level (avgpool)	V1–V4	127.5	247.5	0.281
High-level (avgpool)	V4–IT	132.5	197.5	0.391

D Variable glossary

Symbols used in the main text and this supplement. "Sec." refers to the Methods subsections of the main paper. Information is in nats and time in ms.

Symbol	Meaning	Sec.
$r \in \{\mathrm{V1}, \mathrm{V4}, \mathrm{IT}\}$	Cortical area (region of interest, ROI)	2.1
N	Number of trials in an analysis ($\geq$22,000 in the main analyses). Subsamples: 5,000 trials (bias correction; each window/stride configuration), 500 trials per RSA subsample, 3,000 (GCMI check); 6,000 images in the model-internal analyses	2.1, 2.4, 2.6
k	Trial index in Sec. 2.1 (not the dimensionality $k_r(t)$ below)	2.1
C_r	Number of recording channels in ROI r (Monkey N: 448 V1, 256 V4, 256 IT)	2.1
t	Time-bin index; the bin covers $[t,\ t + W)$; 57 bins with centres 7.5–292.5 ms relative to stimulus onset	2.1
W, s	Window length and stride (15 and 5 ms in the main analyses; robustness sweep over $(W, s) \in \{(10, 2), (15, 4), (20, 5), (30, 5), (50, 10)\}$ ms)	2.1, 2.6
$\mathbf{X}_r(t) \in \mathbb{R}^{N \times C_r}$	Column-centred, window-averaged z-scored MUA of ROI r in bin t	2.3

Data and windowing.

Symbol	Meaning	Sec.
$\hat{\Sigma}_r$, λ_i	Sample covariance of $\mathbf{X}_r(t)$ and its eigenvalues	2.3
γ, γ_{joint}	Aspect ratios: $\gamma = C_r/N$ for a single area; $\gamma_{\mathrm{joint}} = \sum_r k_r(t)/N$ for the second-level pass	2.3
$\hat{\sigma}^2$	Noise variance, estimated as the median positive eigenvalue	2.3
$\hat{\lambda}^+$	Marchenko–Pastur upper edge, $\hat{\sigma}^2(1 + \sqrt{\gamma})^2$; eigenvalues above it are retained as signal	2.3
$k_r(t)$	Signal dimensionality of ROI r in bin t (at least 1); fixed to 60 in the dimensionality-matched control	2.3
$\mathbf{V}_r^{(k)}$	Top $k_r(t)$ eigenvectors of $\hat{\Sigma}_r$	2.3
$\mathbf{S}_r(t)$	Signal-subspace projection of ROI r ($N \times k_r(t)$); a PID source	2.3
$\mathbf{Z}(t)$	Joint representation: concatenated $\{\mathbf{S}_r(t)\}$ after a second MP-PCA pass. $\mathbf{Z}_{\mathrm{all}}$: three areas; $\mathbf{Z}_{r_A r_B}$: an area pair; blocks are built the same way from a pair's projections. Written Z_{joint} in the GCMI section above	2.3, 2.5
$\mathbf{Y}^{(\ell)} \in \mathbb{R}^{N \times 50}$	ResNet-18 target for layer $\ell \in \{$layer1–layer4, avgpool$\}$ (50 PCs); $\mathbf{Y}$ denotes a generic target	2.2

Signal subspace and joint representations.

Symbol	Meaning	Sec.
d_S, d_Y	Dimensionality of $\mathbf{S}$ and $\mathbf{Y}$	2.4
$\hat{\Sigma}_S$, $\hat{\Sigma}_Y$, $\hat{\Sigma}_{SY}$	Sample covariances of $\mathbf{S}$, $\mathbf{Y}$ and $[\mathbf{S}, \mathbf{Y}]$	2.4
$\hat{I}(\mathbf{S}; \mathbf{Y})$	Gaussian plug-in mutual information (linearly accessible information); clipped to zero if a determinant sign is non-positive	2.4
B, N_{sub}	Number of shuffles (500) and trials in the subsample (5,000) used for bias estimation	2.4
$\hat{b}$	Shuffle bias: mean MI over B row-permutations of $\mathbf{Y}$; $\hat{b}_r$ for single areas, $\hat{b}_{\mathbf{Z}}$ for a joint representation	2.4
$\hat{I}^{\mathrm{corr}}$	$\max(0, \hat{I} - \hat{b})$; PID time series are additionally baseline-subtracted (pre-stimulus mean)	2.4

Mutual information and bias correction.

Symbol	Meaning	Sec.
n	Number of neural sources in a PID (three-source: 3; pairwise and block: 2)	2.5
R	Redundancy (minimum mutual information): $\min_r \hat{I}^{\mathrm{corr}}(\mathbf{S}_r; \mathbf{Y})$; "Red" in the GCMI section and Table S4	2.5
U_r	Unique information of source r: $\max(0, \hat{I}^{\mathrm{corr}}(\mathbf{S}_r; \mathbf{Y}) - R)$; "Uniq$_s$" in the GCMI section; U_{V1}, U_{V4}, U_{IT} in Table S4	2.5
Syn	Synergy: $\max(0, \hat{I}^{\mathrm{corr}}(\mathbf{Z}; \mathbf{Y}) - R - \sum_r U_r)$	2.5
I_{joint}, $I(S; \mathbf{R})$	Total joint MI between the joint population and the target (Tables S2, S4)	—
(r_A, r_B)	Area pair in the pairwise PID: V1–IT, V1–V4, V4–IT	2.5
V1+V4 vs. IT	Block PID: the pair's projections are merged by a second MP-PCA pass and treated as one source against the singleton area	2.5
$\mu_{\mathrm{null}}(t)$, $\sigma_{\mathrm{null}}(t)$	Mean and standard deviation of the shuffle null in bin t; the cluster-forming threshold is $\mu_{\mathrm{null}} + 2\sigma_{\mathrm{null}}$	2.6
α	Significance level, 0.05	2.6
Onset	First significant cluster bin in the main analyses; first nonzero bin where a table says so (Tables S8)	2.6
Δ	IT-unique onset minus synergy onset (Table S3)	—
PR	Participation ratio, $(\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$ over an area's covariance eigenspectrum	2.6
Quench %	Percentage decrease of PR from its baseline (mean over 0–100 ms) to its post-stimulus minimum (Table S1)	—

Partial information decomposition, statistics and timing.

Symbol	Meaning
$\tilde{X}_{ij}$, r_{ij}, Φ^{-1}	Copula transform: r_{ij} is the rank of X_{ij} within column j (not the ROI index r), Φ^{-1} the standard normal quantile function, $\tilde{X}_{ij}$ the transformed value
n (copula section)	Number of samples in the copula transform (not the number of PID sources)
$s \in \{V1, IT, V4\}$	Source index in the GCMI PID; corresponds to r in the main text
t1–t4, "final"	CorNet-S recurrent timesteps (V4 has four, IT has two); "final" is the last one. Not the time-bin index t
Marginal balance	Table S7: how similar the two sources' marginal information about the target is; 1 is balanced, values near 0 mean one source dominates

Notation specific to this supplement.

Term	Meaning
MUA	Multi-unit activity
TVSD	THINGS Ventral Stream Spiking Dataset
PID	Partial information decomposition
MMI	Minimum mutual information (redundancy definition)
MP-PCA	Marchenko–Pastur PCA (signal-subspace estimation)
GCMI	Gaussian copula mutual information
RSA, RDM	Representational similarity analysis; representational dissimilarity matrix
PR	Participation ratio

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: To the best of our knowledge, claims are thoroughly segmented and elaborated in the results

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please find the limitations section in the main paper.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: To the best of our knowledge , our findings are purely empirical in nature.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Please find that the Methods section provides a list of methods , the supplementary section provides the compute resources estimated.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Please note that the Dataset used in this paper is a public dataset which belongs to a third party.The code used to reproduce the experiments is included with the submission.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Please find the above in the Methods section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Please find that we have performed a Cluster mass significance test and omnibus test , large sample set of 25000 samples.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Please find the above in the Supplementary section

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have read the code of ethics and abide by it .

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: foundational basic science

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: none

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Please find that the the origin of the dataset and methods have been clearly cited in the main paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification:Please find that only existing methods have been implemented , claims are empirical.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Please note that we did not use crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification:none

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification:Drafting sections of the paper, Editing (e.g., grammar, spelling, word choice), Implementing standard methods

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.