Systems/Circuits

Decomposing Neural Circuit Function into Information Processing Primitives

Nicole Voges,[1,2*] **Vinicius Lima,**[3*] **Johannes Hausmann,**[4] **Andrea Brovelli,**[1,2#] **and Demian Battaglia**[2,3,5#]

[1]Institut de Neurosciences de La Timone, UMR 7289, CNRS, Université Aix-Marseille, Marseille 13005, France, [2]Institute for Language, Communication and the Brain (ILCB), Aix-Marseille Université, Marseille 13005, France, [3]Institut de Neurosciences des Systèmes (INS), UMR 1106, Aix-Marseille Université, Marseille 13005, France, [4]R&D Department, Hyland Switzerland Sarl, Corcelles NE 2035, Switzerland, and [5]University of Strasbourg Institute for Advanced Studies (USIAS), Strasbourg 67000, France

It is challenging to measure how specific aspects of coordinated neural dynamics translate into operations of information processing and, ultimately, cognitive functions. An obstacle is that simple circuit mechanisms—such as self-sustained or propagating activity and nonlinear summation of inputs—do not directly give rise to high-level functions. Nevertheless, they already implement simple the information carried by neural activity. Here, we propose that distinct functions, such as stimulus representation, working memory, or selective attention, stem from different combinations and types of low-level manipulations of information or information processing primitives. To test this hypothesis, we combine approaches from information theory with simulations of multi-scale neural circuits involving interacting brain regions that emulate well-defined cognitive functions. Specifically, we track the information dynamics emergent from patterns of neural dynamics, using quantitative metrics to detect where and when information is actively buffered, transferred or nonlinearly merged, as possible modes of low-level processing (storage, transfer and modification). We find that neuronal subsets maintaining representations in working memory or performing attentional gain modulation are signaled by their boosted involvement in operations of information storage or modification, respectively. Thus, information dynamic metrics, beyond detecting which network units participate in cognitive processing, also promise to specify how and when they do it, that is, through which type of primitive computation, a capability that may be exploited for the analysis of experimental recordings.

Key words: functional connectivity; information processing; information theory; neural circuit modeling; selective attention; working memory

Significance Statement

We can easily name brain functions, and we are well informed about brain structure. However, it is not easy to bridge the gap between the two. Part of the problem is that simple circuit mechanisms do not directly give rise to high-level functions. Yet, they already implement simpler forms of information processing, a sort of "neural assembly language." Here, we track such primitive operations of processing using metrics from information theory and benchmarking them on functional simulations. We thus prove that these metrics can reveal the different flavors of information processing involved in different well-defined functions (working memory, selective attention, etc.). We thus transform descriptions of neuronal dynamics into descriptions of how these dynamics specifically propagate and modify information.

Received Jan. 27, 2023; revised Sept. 1, 2023; accepted Sept. 19, 2023.

Author contributions: N.V., A.B., and D.B. designed research; N.V. and V.L. performed research; N.V. and V.L. analyzed data; V.L., J.H., A.B., and D.B. contributed unpublished reagents/analytic tools; N.V., V.L., A.B., and D.B. wrote the paper.

This work, carried out within the Institute of Convergence ILCB, was supported by grants from France 2030 (ANR-16-CONV-0002) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). A.B. was supported by the PRC project "CausaL" (ANR-18-CE28-0016) and received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). V.L. and D.B. are supported by the Maria Skłodowska-Curie action program of the European Union under the Innovative Training Network "iConn" (H2020 ITN 859937). D.B. was supported by the University of Strasbourg Institute for Advanced Studies (USIAS-2020-044), where part of this research has been conducted. We would like to thank Etienne Combrisson (INT, Marseille) for implementing the functions used to perform the IPP analysis and documentation in the FRITES package. We are also grateful for the inspiring discussions with Joe Lizier, Michael Wibral, and Stefano Panzeri.

*N.V. and V.L. have shared first authorship.

#A.B. and D.B. have shared last and corresponding authorship.

The authors declare no competing financial interest. The research reported here has no relation with the economic activities of the company Hyland Switzerland Sarl, for which J.H. is currently working.

Correspondence should be addressed to Andrea Brovelli at andrea.brovelli@univ-amu.fr or Demian Battaglia at demian.battaglia@univ-amu.fr.

https://doi.org/10.1523/JNEUROSCI.0157-23.2023

Introduction

In a seminal proposal, David Marr observed that a neural system can be described at three different levels: (1) the *functional level* "at which the nature of a computation is expressed," (2) the *algorithmic level* "at which the algorithms implementing a computation are characterized," (3) the *structural level* "at which the mechanisms are realized in hardware" (Marr and Poggio, 1976; see Fig. 1*A* for graphical cartoons of these levels). The third level is directly accessible to experimental investigation: the *structure* of neural circuits (Helmstaedter et al., 2013; Markov et al., 2014), and their activity has been measured across different scales with a variety of techniques. Likewise, one can identify the resulting *function*, such as sensory representation, working memory, or selective attention, and measure the associated cognitive and behavioral performance. However, the quantitative definition of the *algorithmic* level still poses challenges. The assumption is that it comprises information processing operations, which are intermediate steps to produce meaningful functional computations.

Various attempts have been made to identify such canonic computations. Some remain closely inspired by the structural level, emphasizing the role of connectivity motifs (Carandini and Heeger, 2011; Miller, 2016). Others propose to decompose cognitive processes into simpler constituents shared across multiple functions (Taatgen, 2013; Saban et al., 2021). Here, we propose to directly track how patterns of neural activity give rise to emergent transformations of information. Information theory provides the tools to quantify the amount of information in a set of observed signals (Shannon, 1948). Beyond that, recent developments in the frameworks of information dynamics (Lizier, 2013) and partial information decomposition (PID; Williams and Beer, 2010; Wibral et al., 2017) permit the assessment of how such information is propagated and transformed. We define a (nonexhaustive) list of low-level operations —"buffering," "transferring," and "integrating"—which highlights different ways to handle raw information, defined here as *information processing primitives* (IPPs; see Fig. 1*B* for cartoons of a proposed set of primitive operations). We then present a rich toolset of information-theoretical metrics to quantify the enactment of specific IPPs (see Fig. 1*C* for a list of these metrics, matching the IPPs of Fig. 1*B*): we evaluate these metrics on simulated time series produced during the execution of tasks probing certain functions and measure which IPPs are required for their implementation, that is, an *algorithmic decomposition* of the task.

The advantage of data simulated with computational models is arbitrarily many trials in fully controlled dynamical conditions without fluctuating brain states (Shine et al., 2016; Grossman et al., 2019) and a controllable noise level. Our models of choice are neural circuits composed of coupled ring networks. Multi-ring circuits with a known architecture (third Marr's level) can be tailored to implement specific goal function (first Marr's level), so that we can study the IPPs enabled by their dynamics (second Marr's level). Although stylized, they retain important features of cortical connectivity, such as spatial modulation of excitatory–inhibitory recurrent interactions. Ring networks were introduced to study feature-selective representations (Ben-Yishai et al., 1995) and produce a rich spatiotemporal

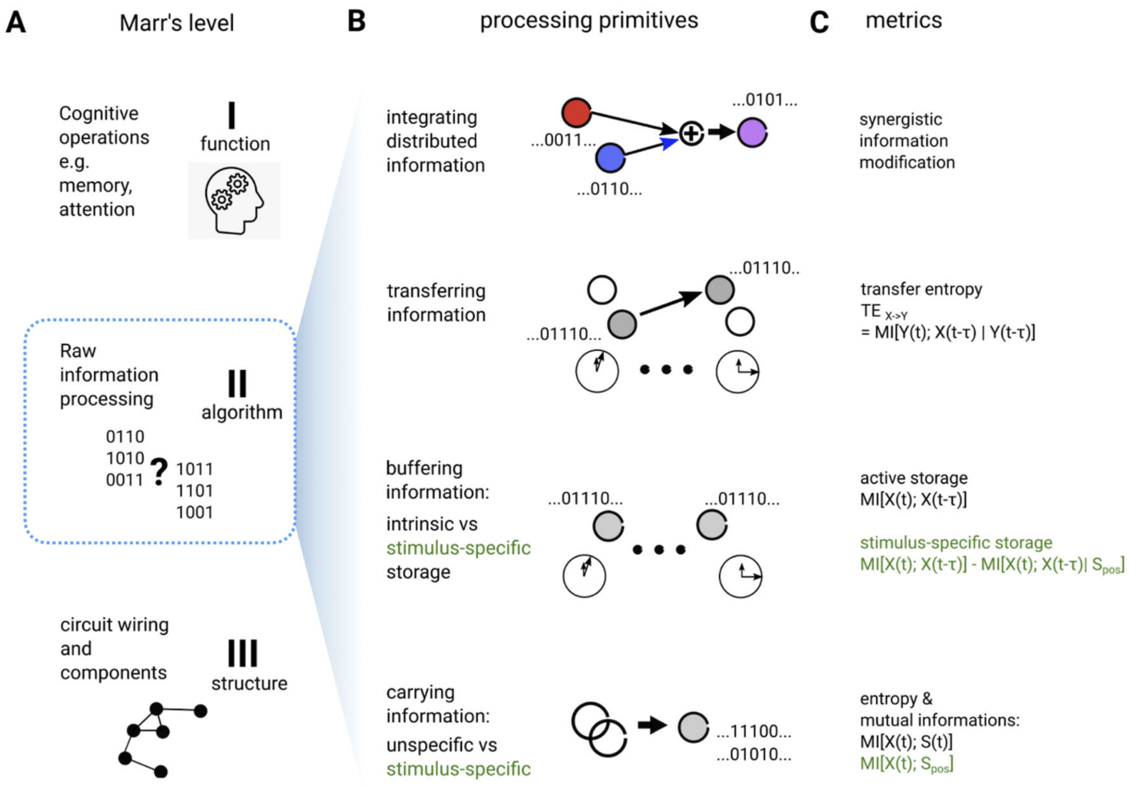

Figure 1. Notions of algorithmic level and information processing primitive operations. ***A***, Neural circuits can be analyzed at three different levels (Marr and Poggio, 1976), here presented from top to bottom: the high-level function performed by the circuit (i.e., the final cognitive operation, first *functional level*, top); the nature of the circuit components (neuronal types, etc.) and the anatomical wiring between them (third *structural level*, bottom); and the second *algorithmic level* of the raw information processing, bridging between circuit structure and function (middle). ***B***, IPPs are elementary algorithmic-level operations performed on streams of information conveyed by neuronal activity, which are involved in building up different functions. IPP's complexity increases from bottom to top. ***C***, The occurrence of such IPPs can be directly tracked and quantified from neural activity data with the corresponding suitable information-theoretical functionals, the metrics used in this paper.

Table 1. List of abbreviations

Abbreviation	Full name	Explanation
SU	Stationary uniform	Dynamical state with spatially homogeneous activity
SB	Stationary bump	Dynamical state with spatially inhomogeneous activity
3FF	Three feed-forward coupled rings	Setup to demonstrate information transfer
2RC	Two reciprocally coupled rings	Setup to demonstrate information integration
FR	Firing rates	Ring model output
MI	Mutual information	Information-theoretical measure
TE	Transfer entropy	Information-theoretical measure
H	Entropy	Information-theoretical measure
GCMI	Gaussian copula mutual information	Alternate method to calculate MI, TE, and H
att-ON	Attention-ON state	Configuration of the 2RC setup
att-OFF	Attention-OFF state	Configuration of the 2RC setup
GBA	Global balanced amplification	Increase in long-range excitation and local inhibition
FLN	Fraction of labeled neurons	Fraction of neurons labeled by the tracer between a given source and target area.

variety of dynamic patterns (Roxin et al., 2005, 2006). Multiple rings can be coupled to account for interactions between multiple cortical layers, columns (Stetter et al., 2000; Battaglia and Hansel, 2011), or even brain regions. Ardid et al. (2007, 2010), for instance, used a network composed of two coupled rings, representing sensory and frontal cortical modules, to model the attentional modulation of responses to oriented visual stimuli. We capitalize on these studies to perform simulated functions: (1) the generation of sensory responses and their maintenance in working memory; (2) the propagation of sensory responses across cortical regions; and (3) the selective attentional modulation of these responses as an effect of top-down influences.

To validate our IPP approach beyond generic and ad hoc constrained ring models, we additionally analyze stimulus-evoked activity in a large-scale, realistic connectome-based model (Joglekar et al., 2018) reproducing some of the circuit mechanisms seen in the ring models (and the associated algorithmic effects).

We provide a proof-of-concept demonstrating algorithmic decompositions on controllable dynamics with well-identified functions. We demonstrate that data-driven information-theoretical metrics are suitable to capture IPPs involved in different functions and to identify which circuit units are participating in specific types of information processing at different spatial positions or times. The proposed framework may thus be suitable for probing the inner workings of actual cognitive processes in electrophysiological or neuroimaging recordings.

Models and Methods

We begin with a description of the basic computational model used in this study to reproduce tuned sensory responses, working memory, signal propagation through a hierarchy of areas, and the modulatory effects of selective attention on sensory responses. Then, we specify three setups demonstrating different types of low-level information processing underlying emulated functional computations. Finally, we detail the IPPs (low-level IPPs) that characterize different types of basic information transformations. A graphical summary of the IPPs at the functional level and the corresponding metrics is given in Figure 1, *B* and *C*. All abbreviations are summarized in Table 1, and all model parameters for the circuit module are presented in Figure 2, and the different numerical experiments performed in Figures 3–5 are specified in Table 2. Table 3 gives the list of regions included in the large-scale model, and Table 4 the parameters of the connectome-based simulations.

Computational model of one region: ring network model

The building block of all ring networks studied here is the rate model version of the one-dimensional ring network with delayed interactions analyzed by Roxin et al. (2005, 2006); see Figure 2*A*. The one-ring network provides a canonical model for a feature-selective cortical module (e.g., a visual cortex hypercolumn). In the ring rate models, the activity of N coupled nodes (also called units) is characterized by their firing rates $R_k(t)$, $k = 0 \dots N{-}1$. The total input to each node k is a linear combination of the activity of all presynaptic nodes, an external stimulus, and the average value of the external drive I_{ext}, indicated In Table 2, is set to have a baseline stationary rate equal to $R_k(t) = \sim 0.1$. In addition, each node k also receives a noisy component $\eta_k(t)$ drawn independently at every time and for every node from the uniform distribution over the interval $[-0.5, 0.5]\ I_{\text{ext}}$.

Each node may receive additional external input I_{stim} at certain times, associated with the presentation of an external stimulus. Such stimulus current is spatially localized to model the nodes' stimulus selectivity. We choose a Gaussian kernel (compare Fig. 2*B*) of prescribed maximum amplitude A_{stim} and width σ_{stim}, centered at a position $S_{\text{pos}} \mathrel{\widehat{=}} \theta_{\text{stim}}$, which varies across different simulated trials. The stimulus time course is given by a function $S(t)$, equal to one during stimulus presentation and zero otherwise.

The time evolution of the activity of each node is governed by a first-order delay differential equation with time delay D involving a threshold-linear input–output transfer function $\Phi(x) = x$ if $x > 0$ and $\Phi(x) = 0$ otherwise:

$$\frac{\mathrm{d}R_k}{\mathrm{d}t} = -R_k(t) + \Phi\left(I_{\text{ext}} + \eta_k(t) + I_{\text{stim}}(k\ t) + \sum_{l \neq k} J_{\text{kl}} R_l(t - D) \right).$$

We consider rings of $N = 100$ nodes, where each node k is labeled by its angular position on the ring $\theta_k = 2\pi k/N$, $k = 0, 1 \dots N{-}1$, coupled to all other nodes $l \neq k$ through a distance-dependent coupling kernel J_{kl}, depending on the angular distance between nodes:

$$J_{\text{kl}} = J_0 + J_1 \cdot \cos(\Delta\theta_{\text{kl}}) \quad \text{with } \Delta\theta_{\text{kl}} = 2\pi(k - l)/N$$

for the link between nodes k and l. The coefficients J_0 and J_1 control the spatial modulation and the net sign of interactions (excitatory or inhibitory) between nodes. Figure 2*C* shows an example of a coupling kernel for the parameters $J_0 = 0$ and $J_1 = 1$, resulting in excitatory short-range interactions with nodes within a range $-\pi/2 \le \Delta\theta_{\text{kl}} \le \pi/2$, and inhibitory long-range (lr) interactions with nodes farther away, that is, $|\Delta\theta_{\text{kl}}| > \pi/2$ ("Mexican hat" profile).

Dynamical regimes of the ring model and properties of stimulus response

The ring network exhibits a rich spectrum of dynamical states, depending on J_0 and J_1. Figure 2*D* (top) shows a schematic phase diagram. The model exhibits a stationary fixed-point solution for small coupling values J_0 and J_1, corresponding to an asynchronous regime in which the average firing rate is constant in time and spatially homogeneous in the absence of external stimuli [stationary uniform (SU) regime]. When modifying J_0 and J_1, the SU regime loses stability. For large J_0 or J_1, the firing rate of every node explodes toward infinitely large values, as the chosen

Table 2. List of parameters for ring model simulations

Parameter	Description	Values	Remark
δt_{int}	Integration time step	0.01	Arbitrary units
δt	Time step for the analysis	10 δt_{int}	
D	Delay between ring units	1 δt	
D_{lr}	External delay between rings	2 δt	
dt	Time delay for MI analysis	40 δt	For binning
N	Number of units	100	0 to 99
v	Noise	50%	Relative to the external drive
J_0 for SU state	Internal coupling	−30	SU activity
J_1 for SU state	Internal coupling	−8	SU activity
J_0 for SB state	Internal coupling	−25	SB activity
J_1 for SB state	Internal coupling	11	SB activity
A_{lr}	External forward coupling strength	35	for 3FF rings
σ_{lr}	External forward coupling width	3 nodes	For 3FF rings
A_{lr}	External forward coupling strength	15	For 2RC rings
σ_{lr}	External forward coupling width	3 nodes	For 2RC rings
A_{lr}	External backward coupling strength	23	For 2RC rings
σ_{lr}	External backward coupling width	6 nodes	For 2RC rings
A_{stim}	Constant stimulus amplitude	2.0	Rate $S(t)$ units
σ_{stim}	Stimulus width	8 nodes	
S_{pos} (4 features)	Stimulus injection position	0, 25, 50, 75	For one ring and 3FF rings
S_{pos} (1 feature)	First stimulus position	50	For 2RC rings
S_{pos2} (10 features)	Second stimulus position	0, 10, … 90	For 2RC rings
t_{ON}	Stimulus onset	100 δt	For one ring and 3FF rings
t_{OFF}	Stimulus offset	250 δt	For one ring and 3FF rings
t_{end}	End of simulation	450 δt	For one ring and 3FF rings
t_{ON}	First stimulus onset	110 δt	For 2RC rings
t_{OFF}	First stimulus offset	310 δt	For 2RC rings
t_{ON2}	Second stimulus onset	460 δt	For 2RC rings
t_{end2}	End of simulation	610 δt	For 2RC rings

Table 3. List of regions included in connectome-based model simulations

Included regions
V1, V2, V4, DP, MT, 8m, 5, 8l, TEO, 2, F1, STPc, 7A, 46d, 10, 9/46v, 9/46d, F5, TEpd, PBr, 7m, 7B, F2, STPi, PROm, F7, 8B, STPr, 24c

Table 4. List of parameters for large-scale connectome-based model simulations

	Weak GBA	Strong GBA
ω_{EI} $[pA/Hz]$	12.5	
ω_{II} $[pA/Hz]$	12.5	
ω_{EE} $[pA/Hz]$	24.3	
μ_{EI} $[pA/Hz]$	24.3	
β_E/β_I	0.066/0.351	
rBG_E/rBG_I	10/35	
$r(0)_E/r(0)_I$	10/35	
τ_E/τ_I	20/10	
η	0.68	
τ_i/τ_f[ms]	200/225	
ω_{IE} $[pA/Hz]$	19.7	25/2
μ_{EE} $[pA/Hz]$	33.7	51.5
$I(V1)_E$ $[pA/Hz]$	41.90	21.93

For details on the mathematical formulation of the model and its parameters, see Joglekar et al., (2018) or the aforementioned ReScience paper (github.com/ViniciusLima94/ReScience-Joglekar).

threshold-linear transfer function does not saturate. When $J_0 < 0$, that is, on average negative collective interactions, there is a finite range of positive J_1 Mexican hat modulation, for which the system's activity spontaneously gives rise to localized bumps of activity. These are centered at some stochastically selected angular position (spontaneous symmetry breaking or "Turing instability") and surrounded by silent nodes [stationary bump (SB) regime]. When the average interaction level becomes strongly inhibitory for $J_0 \ll 0$, the SU regime undergoes a transition ("Hopf instability") to a regime in which the firing rates oscillate homogeneously and in phase. This *oscillatory uniform regime* and other possible regimes (such as traveling waves) are not further explored in this study.

We focus on SU and SB regimes, notably on their responses to externally presented stimuli (Fig. 2*D*, bottom). In the absence of a stimulus, the activation in the SU regime is uniform throughout the network. Upon stimulus presentation at an angle θ_{stim}, a bump of stronger activity centered on θ_{stim} develops due to the locally increased excitatory drive. This silences the surrounding nodes outside of the bump via lateral inhibition. Such a bump can be seen as a representation of the presented stimulus, as its position along the ring follows the stimulus' angle. Once the stimulus ends, that is, the additional I_{stim} input goes back to zero, the bumps dissolve, and activity relaxes back to uniform (fading encoding, Fig. 2*D*, bottom left).

The situation is different in the SB regime where a spontaneously generated bump is already present before the stimulus occurs. In SB, the effect of presenting an externally oriented stimulus is the displacement of the previously existing bump, moving it to the location corresponding to the stimulus angle θ_{stim}. Once the stimulus is removed, the evoked bump continues to exist, because it is self-sustained by local recurrent excitation. It persists for a certain time (persistent encoding, Fig. 2*D*, bottom right), before noise may cause it to drift. In the following, we perform simulations at selected working points within the SU and the SB regimes (see Table 2 for our parameters). Note that the parameters A_{stim} and σ_{stim} of the stimulus are chosen and tuned in a way that the bump evoked by a stimulus in said SU working point has a similar width and amplitude to the bump arising in the SB working point, such that the results obtained from simulations in the two regimes are comparable.

Multiregional architectures: coupled ring networks

To model circuits involving more cortical modules, we use networks composed of multiple coupled rings. Each ring is modeled as the previously described single-ring architecture with the possibility to tune different rings to different dynamical regimes and working points. However, an additional current term I_{lr} must be fed to the transfer function Φ of each unit R_{k} to account for an additional drive provided by the lr coupling to nodes in remote rings:

$$I_{\mathrm{lr}}(k\ \ t) \ = \sum_{q\,\in\,\text{remote ring}} W_{\mathrm{qk}} R_q(t - D_{\mathrm{lr}})$$

where the lr connectivity kernel is W_{qk} is a Gaussian with amplitude A_{lr} and width σ_{lr} centered on k, thus ensuring a strict spatiotopy of interregional excitatory projections. Inter-ring interactions are also delayed and can have an independently tuned, longer delay D_{lr}. We consider two types of multi-ring architectures.

Three-ring network with feed-forward coupling. We study the propagation of a stimulus representation through a hierarchy of different ring modules (Fig. 4). To emulate the transfer of information from one area to another, we couple three rings as a feed-forward chain (Fig. 4*A*), called "3FF rings" setup (compare Tables 1, 2). The bottom ring (R1) represents a sensory cortical area that receives subcortical stimulus-related input. The middle ring R2 receives lr feed-forward input from R1, and the top ring R3 from R2. The couplings from R1 to R2 and R2 to R3 have identical strengths and widths. There is no feedback coupling in order to study the capacity of information-theoretical metrics to capture the processing operation of propagation and transfer through a directed hierarchy.

Reciprocally coupled two-ring network. In a second circuit configuration, two regions simultaneously interact via both feed-forward and feedback connections. The goal is to study the capacity of information-theoretical metrics to track the effects of interacting bottom-up and top-down inputs as occurring, for example, in selective attention (Fig. 5). In this setup, called 2RC rings (compare Tables 1, 2), two rings are reciprocally coupled, similar to the model by Ardid et al. (2007). The bottom ring R1 again constitutes a sensory cortical area, while the top ring R2 represents a prefrontal cortical area. The latter implements working memory, later acting as a source of top-down influences. The parameters of feed-forward and feedback connections are fine-tuned (Table 2) to obtain attention-like enhancements of stimulus response.

Task simulations with the ring models

For the one-ring configuration of Figure 3 and for the 3FF ring setup of Figure 4, we generate 1,000 trials with different noise realizations per each of four possible orientations S_{pos} of the stimulus, thus 4,000 trials in total, for both the SU and SB working points. The stimulus injection center positions S_{pos} are equally spaced along the ring (at angles 0, $\pi/2$, π, and $3\pi/2$), alternating randomly from trial to trial. Time is measured in arbitrary units δt (10 numeric integration steps per δt, fourth-order Runge–Kutta integration scheme, augmented with delay). For all analyses, we drop an initial period of 400 δt to discard early transients. In each simulated trial, we first record 100 δt of baseline dynamics, before injecting a stimulus, which is then maintained for 150 δt, that is, $S(t) = 1$ for $t_{\mathrm{ON}} = 100\ \delta t < t < t_{\mathrm{OFF}} = 250\ \delta t$, and $S(t) = 0$ otherwise.

For the analyses shown in Figure 5 (probing attentional modulation), the task organization is more complex with two stimulus presentations: the first stimulation at position S_{pos} ("cue") starts at $t_{\mathrm{ON}} = 110\ \delta t$ and stops at $t_{\mathrm{OFF}} = 310\ \delta t$; the second stimulation at position S_{pos2} ("match") starts at $t_{\mathrm{ON2}} = 460\ \delta t$ and stops at $t_{\mathrm{OFF2}} = 610\ \delta t$. The cue position is fixed in all trials at $S_{\mathrm{pos}} = \pi$, while S_{pos2} alternates randomly between 0 and 2π, in steps of $2\pi/10$. These stimulus combinations are generated for two different conditions. The first condition is called "attention-OFF" (att-OFF), mimicking empirical experiments in which no attentional modulations are expected [as the receptive field of recorded neurons is not attended to, i.e., the "unattended" condition described by Treue (2001)]. In att-OFF, both the bottom ring R1 and the top ring R2 of the 2RC setup are in SU state; thus, bump representations of cue and match stimuli are formed during stimulus presentation and decay thereafter. The second condition is called "attention-ON" (att-ON), mimicking experiments in which attentional modulations of responses *are* expected [as attention is directed to the receptive field of the recorded neurons, i.e., the "attended" condition described by Treue (2001) and Maunsell and Treue (2006)]. The information that attention must be engaged toward the cue is provided shortly before cue presentation: at time $t_{\mathrm{switch}} = 100\ \delta t$, the top ring R2 is moved from the SU to an SB regime (parameters are detailed in Table 2). As a result of attention being "switched on," the top ring R2 maintains a persistent representation of the cue through the entire delay period between the offset of the cue at t_{OFF} and the onset of the match stimulus at t_{ON2}. This working memory representation interacts nonlinearly with the representation evoked by the match in R1. To generate Figure 5, we run 5,000 trials for each S_{pos} and S_{pos2} combination, in both att-ON and att-OFF conditions.

Estimating information-theoretical quantities

We track the effects of neural function at the algorithmic level by quantifying how simulated dynamic patterns translate into elementary information transformations. All information-theoretical metrics we introduce to detect the enactment of different IPPs can be seen as elaborations of a few basic quantities (Cover and Thomas, 2006). The amount of information carried on average by observations of a random variable X is quantified by Shannon Entropy:

$$H(X) \ = \ -\sum_{x \in X} P(x) \log_2 P(x)$$

which is a functional of the empirical probabilities $P(X)$ of observing different possible values of the variable X. Conditional entropy quantifies the amount of information needed to describe the outcome of a random variable X given that the value of another random variable Y is known, it is defined as follows:

$$H(X|Y) \ = \ -\sum_{x \in X y \in Y} P(x, y) \log_2 P(x \mid y).$$

The mutual information (MI) between X and Y quantifies the statistical dependence between the two variables, and it is defined as the difference between marginal and conditional entropies:

$$\mathrm{MI}(X;\, Y) = H(X) - H(X|Y) = \sum_{x \in X y \in Y} P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)}.$$

It describes the fraction of information, which is shared, that is, redundantly encoded by both X and Y.

The conditional MI between X and Y conditioned on a third variable Z is defined as follows:

$$\mathrm{MI}(X;\, Y|Z) = \sum_{x \in X y \in Y z \in Z} P(x, y, z) \log_2 \frac{P(x, y \mid z)}{P(x \mid z)P(y \mid z)}$$

providing the average amount of information redundantly carried by X and Y, which is not already carried by Z.

A crucial step in evaluating any information-theoretical quantity is the proper estimation of the empirical probability distributions of one or more observables jointly. For most analyses of the results obtained with the (coupled) ring models, we use "plug-in" or "direct" estimators, biased for the small amounts of data typically available in neurophysiological experiments but converging to stable values for large datasets. We estimate histograms of firing rate variables using 24 equally spaced bins (qualitatively analogous results are obtained using 18 and 32 bins).

For some other analyses (Fig. 6 and Fig. 4-1*C*), estimates are computed using a semi-parametric Gaussian copula approach (Ince et al., 2017). The Gaussian copula mutual information (GCMI) approach exploits the fact that MI is invariant under monotonic transformations of the marginals. This can be exploited to render the joint distribution of Gaussian variables by means of local transformations on the marginals, using the so-called Gaussian copula. It involves calculating the inverse standard normal cumulative density function (CDF) of the empirical CDF of each sample, separately for each input dimension. Then, entropy values can be estimated using a standard covariance-based formula for Gaussian variables. It includes a parametric bias correction for estimates and an analytic correction to compensate for the bias due to the estimation of the covariance matrix from limited data. GCMI is a robust rank-based approach that allows the detection of any type of relation as long as this relation is monotone. GCMI is computed using functions implemented in the Frites Python toolbox (Combrisson et al., 2022a,b). Finally, information-theoretical quantities are normalized by the largest of the entropies of the involved variables [e.g., normalizing MI(X; Y) to MI(X; Y)/max(H(X(H(Y))], so that they are bounded in the unit interval and express relative fractions of information rather than absolute amounts.

We now present in detail the specific information-theoretical quantities that we use to define and track IPPs.

Tracking IPPs

We focus on a simple set of elementary processes of information manipulation stemming from neural activity. To detect their presence and quantify the degree to which different circuit nodes are engaged at different times along the simulated tasks, we use the above information-theoretical functionals. The simplest possible operation one can perform with information is to *carry* it. Once a network node carries some information it can keep carrying it actively for an extended time, that is, it can *buffer* it, or it can push it to another network node, that is, *transfer* it. The most complicated primitive operation we consider is *integrating* multiple streams of information, that is, combining information from multiple sources to reveal the existence of information that was inaccessible prior to their combination. Figure 1*B* shows cartoons illustrating these basic IPPs (in ascending order of complexity from bottom to top), and Figure 1*C* lists the associated information-theoretical functionals.

Note that the metrics we use to quantify IPPs have some limitations. For instance, the functional to quantify transfer conflates into the estimated quantity of “transferred” information (in quotes) a part of the information conveyed by nonlinear and nonlocal synergies. We comment on several of these limitations and drawbacks later in the Discussion. For now, be aware that throughout the entire manuscript, terms such as “buffering,” “transferring,” or “integrating” refer to nothing more than the operative definitions provided in the following and may thus not perfectly correspond to other formulations from more elaborate theories.

The IPP of carrying information or local encoding of information: entropy and MI. The average information carried by a network node at time t is given by the functional $H(R(t))$, where the across-trial firing rate $R(t)$ (sampled at time t) *is used to estimate* the probability distribution $P(R(t))$. One can also evaluate the fraction of the information carried relative to the presented stimulus, either by considering stimulus presence or absence MI($R(t)$; $S(t)$), where $S(t)$ is the spatially inhomogeneous stimulus time course, or the feature carried by the stimulus, MI($R(t)$; S_{pos}), where S_{pos} is the stimulus’ orientation angle. Both MI terms are then normalized by $H(R(t))$ to be expressed in relative form.

The IPP of buffering information: active information storage. A circuit node buffering information continues carrying some information, which was already present. This primitive processing operation can be quantified by *active information storage* (AIS; Lizier et al., 2012; Wibral et al., 2014). In its simplest manifestation, it corresponds to the MI between past and present activity, MI($R(t)$; $R(t-\tau)$), where τ is an adjustable latency, set to $\tau = 40\ \delta t$, unless otherwise specified.

Additionally, one may evaluate the fraction of information about the orientation of the stimulus presented in a trial actively buffered by a node. The resulting *stimulus-specific active storage* is given as follows: MI($R(t)$; $R(t$-τ))−MI($R(t)$; $R(t$-$\tau)|S_{\mathrm{pos}}$), that is, the totally stored information minus the part of this stored information, which does *not* depend on the stimulus orientation. This corresponds to the so-called co-information, or negative interaction information, between three variables $R(t)$, $R(t$-$\tau)$, and S_{pos} (McGill, 1954; Ince et al., 2017). Note that co-interaction and interaction information may be negative when there are synergistic interactions involved (see Discussion). However, in our simulations, we always obtain positive values—indicating that the correlation between rates $R(t)$ and $R(t$-$\tau)$ is indeed mostly accounted for by redundant information about S_{pos}.

Finally, in all cases, storage measures are also normalized by $H(R(t$-$\tau))$ to make them relative.

The IPP of transferring information: transfer entropy. Information transfer between neural populations can be estimated from the statistical dependencies between neural signals (Brovelli et al., 2004; Bressler and Seth, 2011) using model-free methods relying on the Wiener–Granger principle (Wiener, 1956; Granger, 1969). It identifies information transfer between time series when future values of a given signal can be predicted from the past values of another signal, above and beyond what can be achieved from its autocorrelation. A general information-theoretical measure to detect directed information transfer according to the Wiener–Granger principle is transfer entropy (TE) (Schreiber, 2000); it captures any (linear and nonlinear) time-lagged conditional dependence between neural signals (Vicente et al., 2011). We remind once again that the “transfer” captured by TE also conflates synergistic information atoms and is thus not the strictest definition of transfer that may be possible (see Discussion). We however favor it here for its simplicity of implementation. TE from X to Y is defined as the conditional MI between the past of X and the present of Y, conditioned on the past of Y:

$$\mathrm{TE}_{X\rightarrow Y} = \mathrm{MI}(Y(t);\ X(t-\tau)|Y(t-\tau)).$$

Note that TE is asymmetric, that is, $\mathrm{TE}_{X\rightarrow Y} \neq \mathrm{TE}_{Y\rightarrow X}$, thus a suitable measure for directed functional connectivity (Battaglia et al., 2012; Palmigiano et al., 2017). We compute two types of TE. First, active transfer from stimulus time course to response rate $\mathrm{TE}_{S\rightarrow R}(t) = \mathrm{MI}(R(t);\ S(t\text{-}\tau)\ |\ R(t\text{-}\tau))$ in the one-ring configuration. Second, active transfer between the activities $\mathrm{R1}_k$ of node k in ring 1 and $\mathrm{R2}_k$ of another homologous node (with identical angular coordinate) located in a second ring $\mathrm{TE}_{\mathrm{R1},k\ \rightarrow\ \mathrm{R2},k}(t) = \mathrm{MI}(\mathrm{R2}_k(t);\ \mathrm{R1}_k(t\text{-}\tau)\ |\ \mathrm{R2}_k(t\text{-}\tau))$ in the 3FF ring setup (Fig. 4). For comparison (and assessment of numerical estimation artifacts), we also compute the backward terms $\mathrm{TE}_{R\rightarrow S}(t)$ and $\mathrm{TE}_{\mathrm{R2}\rightarrow \mathrm{R1},k}(t)$ that should be zero by construction since there are no feedback couplings. Again, TE is normalized by the entropy of the source variable. For the 3FF ring setup in SB state, we additionally calculate $\mathrm{TE}_{\mathrm{RN},k\ \rightarrow\ \mathrm{RN}+1,k}(t)$ and $\mathrm{TE}_{\mathrm{RN}+1,k\rightarrow \mathrm{RN},k}(t)$, with $N = 1, 2$ the ring number, using multiple delays ($dt = 5, 10, 15, \dots\ 40\ \delta\mathrm{t}$) simultaneously without entropy normalization. For this analysis, we use the GCMI approach implemented in the Frites Python toolbox (see the Methods subsection on downloadable codes). Finally, we remark that stimulus-specific analogs of Transfer Entropy exists (Bím et al., 2020; Pica et al., 2019), as for active information storage, however we don't make use of them here.

The IPP of integrating information: synergistic modification. A third type of primitive processing operation can arise when two input sources X_1 and X_2 interact and communicate with a common target Y. Synergy may emerge, where extra information is conveyed by the interaction between the sources. This implies that the combined inputs X_1 and X_2 provide surplus information with respect to the inputs considered separately (Brenner et al., 2000; Latham and Nirenberg, 2005). The process of extracting this surplus information—performed by the output node Y—has been called *synergistic modification* (Lizier et al., 2013, 2018).

Our two source variables are the firing rate of a node in one ring and a stimulus position (X_1 and X_2), while the target variable is the firing rate of a node in a second ring (Y). We suggest that a primitive processing operation is the extraction of synergistic information by the target node. To do so, we exploit a formalism that allows the decomposition of multivariate MI between a system of predictors and a target variable. It quantifies the information that several predictors provide uniquely, redundantly, or synergistically about a target variable, the so-called PID (Williams and Beer, 2010). The PID formalism can be outlined using the information

Venn diagram (Fig. 5*E*). The total information that the output Y carries about the pair of inputs (X_1, X_2) consists of *unique*, *redundant*, and *synergistic* parts. The information Y shares with X_1, but not with X_2, is denoted as $\mathrm{MI}(Y; X_1 \setminus X_2)$, and, conversely, the other unique information term is denoted as $\mathrm{MI}(Y; X_2 \setminus X_1)$. The redundant part is the information shared by both X_1 and X_2 is denoted as $\mathrm{MI}(Y; X_1 \cap X_2)$. The remaining part is thus synergy whose amount can be determined by subtracting the nonsynergistic fractions from the total amount of information $\mathrm{MI}(Y; (X_1, X_2))$ that Y carries about the pair of inputs. However, these quantities cannot be estimated directly. We operate under the so-called minimum mutual information (MMI) ansatz, which has been shown to provide correct estimations for Gaussian systems (Barrett, 2015, Luppi et al., 2022). According to MMI, redundant information can be computed as the minimum of the information provided by each individual source to the target (i.e., an upper bound estimation: one input is supposed not to carry any unique information).

$$\mathrm{Red} = \min\{\mathrm{MI}(X_1; Y)\quad \mathrm{MI}(X_2; Y)\}.$$

Then, the synergistic information can be computed by subtracting the MI of the single source variables from the total information. Each of the individual $\mathrm{MI}(X_1; Y)$ and $\mathrm{MI}(X_2; Y)$ terms is the sum of unique contributions from the considered variable and the redundant fraction. Thus, when subtracting these terms from the total $\mathrm{MI}(X_1, X_{2;} Y)$, we subtract redundancy twice, so we have to "add back" one red term (in order to subtract it only once, as we should). Therefore, with some simple algebra (see the Venn diagram of Fig. 5*E*), we can write the following:

$$\mathrm{Syn} = \mathrm{MI}(X_1\ X_2; Y) - \mathrm{MI}(X_1; Y) - \mathrm{MI}(X_2; Y) + \mathrm{Red}.$$

The two equations represent the redundant and synergistic information carried by the co-modulations in firing rate X_1 and input-related variable X_2 about the target node Y, respectively. Once again, this metric is normalized by an entropy term, here the one of the target output Y, to evaluate the synergistic fraction of the total output information flow.

Large-scale connectome-based model

Model implementation. To go beyond generic models of interacting areas, we also re-implemented a large-scale model of cortical activity previously published by Joglekar et al. (2018). It consists of 29 cortical areas (Table 3) whose local activity is modeled with simple neural mass equations, that is, population dynamics are represented as firing rates. Each local circuit is composed of an excitatory (E) and an inhibitory (I) population with the corresponding EI, IE, EE, and II couplings (Fig. 6*B*, top row). Every region is driven by a baseline current term (representing ongoing background activity). In addition, transient sensory stimulation can be implemented via additional local activity injection (see later).

All areas are reciprocally coupled to each other according to a connectivity matrix derived from systematic tracer experiments in nonhuman primates (Markov et al., 2014; Fig. 6*A*, left). Such a directed and weighted connectome defines the strength of the input each region receives from all other cortical areas and is given by the fraction of labeled neurons (FLN) by the tracer in a given source area, normalized by the total number of labeled neurons over cortical regions. The interareal lr connections are purely excitatory. Different from the original model (Joglekar et al., 2018), we add transmission delays between areas in the model, which have also been estimated empirically (Markov et al., 2013). They rely on the (physical) distance between areas (Fig. 6*A*, right), which is converted to propagation delays by assuming a propagation speed of 3.5 m/s. Cortical areas are ordered according to their hierarchical level (Markov et al., 2013). Following Joglekar et al. (2018), the local circuits in higher-order regions are endowed with stronger excitability and correspondingly longer time constants.

The "bow-tie" architecture of the connectome (Markov et al., 2013, 2014) introduces a barrier to the free propagation of externally injected stimuli. Mimicking a visual stimulus presentation, Figure 6*B* (bottom row) shows the rate dynamics for a brief transient input injection to area V1, which leads to a very limited propagation along the visual stream. To enable a dynamical regime with enhanced signal propagation, Joglekar et al. (2018) introduce a global balanced amplification (GBA) mechanism in which lr excitatory connectivity is strengthened, while simultaneously strengthening the local inhibition (IE connections) within each region (Fig. 6*B*, top row). In the weak GBA regime (left), the stimulus-related activity volley remains local while it reaches higher-order regions due to the facilitated propagation in the strong GBA regime (right). The increased amount of excitation is controlled by a modified inhibitory-to-excitatory balance. In this manuscript, we use these weak and strong GBA regimes as equivalent, respectively, to the "attend-OFF" and "attend-ON" scenarios defined for the coupled ring models.

All parameters for model simulations are listed in Table 4. Full details about the reimplementation of the model by Joglekar et al. (2018) can be found at the following link: github.com/ViniciusLima94/ReScience-Joglekar.

Stimulus presentation and IPP analyses. To investigate the information processing in the large-scale model, we apply a brief stimulus injection of 0.5 ms to the excitatory population in area V1, in the weak and strong GBA regime. Note that for the large-scale model, in contrast to the ring model simulations, there is no stimulus orientation. The stimulation tag S (S_{pos} for the ring models) merely represents the presence or absence of the stimulus. Here, we analyze the IPPs of "buffering," information transfer, synergistic integration, and stimulus-specific storage using the metrics described above. Note that all analyses are based on the GCMI approach as implemented in the Frites Python toolbox (Combrisson et al., 2022b). In the Frites implementation, latency-dependent functionals such as AIS and TE are computed for all latencies τ until a maximum latency and then averaged over this range of τ's (as we did already in supporting analyses of the 3FF ring configuration).

AIS is computed as $\mathrm{MI}(R_n(t); R_n(t\text{-}\tau))$ with a maximum latency τ of 40 ms, where R_n represents the firing rates of the cortical areas $n = 1\ldots N$. In addition, the stimulus-specific AIS is calculated as $\mathrm{MI}(R_n(t); R_n(t\text{-}\tau)) - \mathrm{MI}(R_n(t); R_n(t\text{-}\tau)|S)$, with S indicating stimulus presence or absence. TE from V1 to all other areas is computed as $\mathrm{TE}_{\mathrm{V1}\rightarrow N} = \mathrm{MI}(R_n(t); R_{\mathrm{V1}}(t\text{-}\tau) \mid R_n(t\text{-}\tau))$ with R_{V1} representing the rate in V1 and $n = 1\ldots N - 1$ indexing all other areas. Likewise, we calculate the backward transfer $\mathrm{TE}_{N\rightarrow \mathrm{V1}} = \mathrm{MI}(R_{\mathrm{V1}}(t); R_n(t\text{-}\tau) \mid R_{\mathrm{V1}}(t\text{-}\tau))$ from all other areas to V1. Finally, we compute the synergistic information encoded by V1 ($R_{\mathrm{V1}}(t)$), emerging from the integration of stimulus-related (bottom-up) input S and the feedback (top-down) signals $R_n(t)$ from other cortical areas as $\mathrm{Syn} = \mathrm{MI}(R_n(t), S; R_{\mathrm{V1}}(t)) - \mathrm{MI}(R_n(t); R_{\mathrm{V1}}(t)) - \mathrm{MI}(S; R_{\mathrm{V1}}(t)) + \mathrm{Red}$. To investigate the relation between timing and distance, we calculate the peak latency [time of the synergy peak—time of the rate $R_n(t)$ peaks] and correlate it to the distance of all areas to V1.

Using this model, we simulated 1,000 trials with stimulation and 1,000 without stimulation, each in the weak and strong GBA regimes, respectively, amounting to a total of 4,000 trials. The total simulation time is 7 s, and the initial 4 s is discarded. Stimulus onset occurs at $t =$ 4.5 seconds with a duration of 500 ms. The strength of the stimulus applied to V1 is 42 and 22 pA for weak and strong GBA, respectively. The model is implemented using the NEST simulator toolbox for Python (Eppler et al., 2008).

Available code resources

A python implementation of the ring models (three Jupyter Notebooks) and the large-scale connectome-based model (run "main.py" to generate the simulated data) are available in the GitHub repository: github.com/brainets/IPP_PAPER/tree/main/src. The code for the analyses shown in Figure 6 is also given in a Jupyter Notebook. More instructions can be found in the README file. The connectivity data from Markov et al. (2014) is originally available at the web-site "core-nets.org" and is also given in the following: github.com/brainets/IPP_PAPER/blob/main/src/interareal/markov2014.npy.

The functions used to perform the IPP analysis can be found in the Frites package (Combrisson et al., 2022b). The synergy-based metrics can be computed using the following code: github.com/brainets/frites/blob/master/frites/conn/conn_pid.py.

The metrics based on interaction information (e.g., AIS) can be computed using the following: github.com/brainets/frites/blob/master/frites/conn/conn_ii.py.

Exemplar code and documentation can be found on these web pages: brainets.github.io/frites/auto_examples/conn/plot_pid.html and brainets.github.io/frites/auto_examples/conn/plot_ii.html.

We also have a faster C code for ring model simulations, available on request.

Results

Two dynamic regimes of response to stimulus

As presented in detail in the Models and Methods section, we model a stimulus-selective cortical region as a ring network with a "Mexican hat" connectivity profile (Fig. 2*C*), that is, more excitatory interactions with units (also called nodes) at nearby positions along the ring and more inhibitory with nodes farther away. In our ring model, the activity of coupled network units is characterized by their noisy firing rates. The corresponding phase diagram, that is, its parameter-dependent dynamical regimes, is shown in Figure 2*D*. We measure simulated responses to stimulus presentations (emulating specific tasks) in two different regimes: an SU and an SB dynamic working points (Fig. 2*D*, bottom).

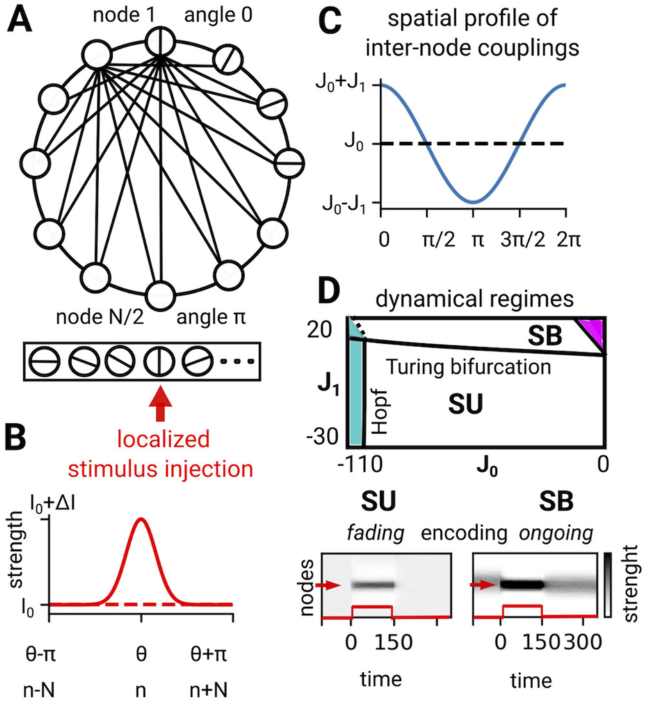

Figure 2. Dynamical states of the ring model of one cortical region. ***A***, Ring model to emulate information encoding and storage: circles represent feature-specific neuronal nodes, parametrized by an angle coordinate θ along the ring, indicating the preferred stimulus direction (denoted by differently oriented lines within the circles); connecting edges indicate internal all-to-all structural couplings, whose weights depend on the distance between the coupled network nodes (compare ***C***). The rectangle below the ring indicates that stimulus-related inputs are injected in a localized fashion to network nodes with a specific stimulus direction preference (indicated by a red arrow), following ***B***, a narrowly tuned Gaussian spatial profile. ***C***, Example profile of spatial modulation for internal ring couplings, here with a "Mexican hat" shape for parameters $J_0 = 0$ and $J_1 = 1$ (see Methods). ***D***, Top: phase diagram reporting different dynamical regimes obtained for different coupling parameter values. Bottom: spatial maps (nodes vs time) for the two dynamical regimes explored in this study: SU activity with transient, stimulus-induced bumps of activity and SB activity with an ongoing, self-sustained bump, persisting after stimulus offset.

The simplest possible task entails (correctly) responding to the presentation of stimuli with different orientations/angles S_{pos} in different trials. Each stimulus is spatially localized, yielding a stimulus selectivity of different units, and can be transiently presented at any chosen time. We show simulated recordings of the activity of units in a single receiving region in Figure 3*A*. The firing rates measured in six exemplary trials are shown in Figure 3*B*, three operating in the SU regime (top) and three in the SB regime (bottom). For each of the trials, we show spatial maps of activity, where the horizontal axis represents time, the vertical axis different units along the ring network, and the firing rate is color-coded. The curves below the spatial maps show the corresponding firing rates for all units over time. We highlight in color the time series of units located at specific positions (indicated by matching-color lines on the corresponding spatial maps). As anticipated, we observe a clear distinction between responses in the SU and SB dynamic working points. In the SU regime, a localized stimulus injection generates a bump-like pattern of increased activity around the injection center (red lines and curves), which disappears soon after the stimulus is switched off. The firing rates of units outside a neighborhood of the injection are either unaffected (blue trace in SU trial #*n* and trial #*j*) or, at larger distances from the injection site, decrease due to the increased lateral inhibition exerted by active units inside the bump (blue trace in trial #*k*). In contrast, in the SB regime, bumps of increased activity develop spontaneously, at random positions along the ring, even without any stimulation. Upon stimulation, the bump's position shifts toward the injection site, its amplitude increases, and it remains stable until the end of the stimulus. After stimulus removal, the bump amplitude relaxes back to its initial value, but the position remains stable around the injection site.

These two configurations correspond to distinct functional behaviors: in SU, stimuli are only *transiently* represented. In contrast, in SB, a sustained activation is observed, supporting the maintenance of the *working memory* of the presented stimulus after its removal. The two IPPs associated with these two simple functional behaviors are "carrying" and "buffering" information, respectively (compare Fig. 1*B*, top two cartoons). They can be tracked by different information-theoretical metrics. We focus first on the "carrying" IPP and address the "buffering" IPP in the next section.

IPP analysis can track the representation of a stimulus

In Figure 3*C* (top), we show the amount of total information that a ring unit carries as a function of time, averaged across all units and trials, provided by the entropy $H(R(t))$ of the activity rates $R(t)$ as they change in response to stimulus presentation. The average entropy of activity differs between the SU and SB regimes. In SU, entropy is rather low in the absence of stimulus, due to the temporal and spatial homogeneity of baseline firing rates, fluctuating only due to a weak background noise. Entropy increases during stimulus presentation, as stimulus-evoked bumps emerge and produce differences in response rates for different stimulus positions, increasing inter-trial variance. In contrast, in SB, entropy is higher before stimulation, because of the stochasticity of the spontaneously emergent bump positions. In our simulated experiments, stimuli are presented at four discrete possible positions (a configuration often met in empirical experiments). Therefore, upon stimulation, the bump positions are quenched to only four pronounced maxima (compare spatial maps in Extended Data Fig. 3-1*A*), resulting in a strong entropy reduction with respect to baseline.

Entropy rises again after stimulus removal, as firing rates are reduced and noisy fluctuations more evident. For both the SU and SB regimes, we furthermore observe tiny peaks and kinks of the entropy time courses in correspondence with stimulus onset and offset at times t_{ON} and t_{OFF}. These variations can be explained by the non-instantaneous response of the system to instantaneous inputs, causing fast transient dynamics (like impulse responses) to occur shortly after stimulus onset and offset (so that inter-trial variance is temporarily increased during these transients).

This simple analysis illustrates how strongly the values of information-theoretical quantities depend on aspects of the neural recordings, such as signal-to-noise ratio and actual task design, which have little to do with algorithmic-level operations. Entropy is an upper bound to other metrics, for example, MI that quantifies the statistical dependence between two or more variables. Entropy variations could result in absolute MI variations, which simply reflect changes in the available informational bandwidth. For the study of primitive processing operations, we thus focus on relative metrics that are normalized by entropy. We therefore do not have to account for the additional complexity of total entropy fluctuations unrelated to the processing probed.

We show the relative amount of information that a unit's activity carries about the stimulus, disentangling two of its aspects: the stimulus time course $S(t)$, that is, its presence/absence at specific times t and the orientation S_{pos} of the stimulus, which is a trial-specific property (changed every trial, see Models and Methods). This fraction of information is captured by the entropy-normalized MI between firing rate $R(t)$ and stimulus time course $S(t)$ (MI($R(t)$; $S(t)$)) in Figure 3*C*, middle, or stimulus orientation label S_{pos} (MI($R(t)$; S_{pos}) in Figure 3*C*, bottom (see Extended Data Fig. 3-1*B*,*C* for the corresponding spatially resolved maps). Before stimulus presentation, no information about the stimulus can be extracted from the neural activity, since the baseline noise entropy is unrelated to the stimulus. In contrast, during stimulus presentation, activity is modulated in dependence on stimulus position (i.e., the distance between S_{pos} and the angle of the recorded unit).

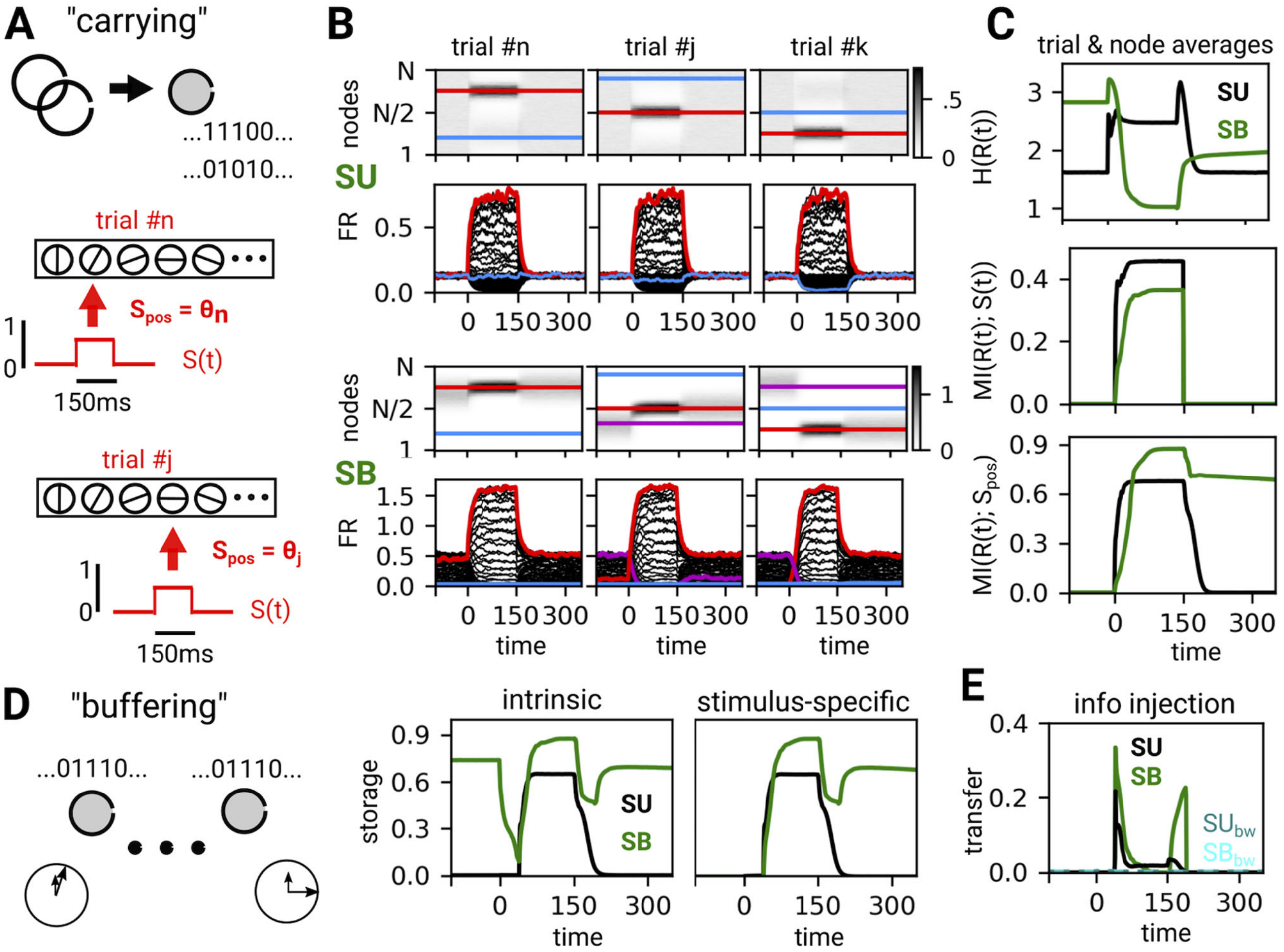

Figure 3. Information encoding and storage in a single-region circuit. ***A***, To track the simplest possible IPP of information "carrying," we simulate different trials in which stimuli with different directions θ are presented for a short, fixed time of 150 ms (as indicated by the stimulus-related input time course $S(t)$). The direction S_{pos} of the presented stimulus is denoted by a red arrow as in Figure 2*A*. ***B***, Spatial maps (top row, units vs time) and single-trial firing rate traces (bottom row) of neuronal activity in a one-ring network, in the SU and SB dynamical regimes. Red lines indicate traces for nodes located at the stimulus center, blue lines nodes far from it, and magenta lines indicate the initial bump position for the SB regime (SB trial #*n*: red covers magenta line). ***C***, Time courses of entropy (top row) and stimulus-related MI averaged over nodes and trials (normalized by entropy), for SU (black) and SB (green lines) regimes. Middle and bottom rows: stimulus presence and orientation (stimulus-specific) are transiently encoded by activity in both SU and SB regimes, as revealed by the MI between rates and, respectively, $S(t)$ and S_{pos}. ***D***, Moving to the IPP of "buffering," we quantify and show time courses of AIS (intrinsic and stimulus-specific) in both the SU and SB regimes averaged over nodes and trials. Stimulus-specific storage persists after stimulus offset in SB, denoting working memory implementation. ***E***, Time courses of information transfer from injected stimulus to rates, quantified by TE. Light and dark cyan lines indicate (negligible) backward transfer from rate to stimulus. See also Extended Data Figure 3-1.

The information about the stimulus-related time course $S(t)$ becomes positive during stimulus presentation, in both the SU and SB regimes, saturating to a higher value in SU (Fig. 3C, middle), as the evoked bump activity configuration differs strongly from the homogeneous baseline. During stimulus presentation, some units develop much lower or higher firing rates than in spontaneous conditions, thus signatures of stimulus presence. No $S(t)$-related information exists before or after stimulus presentation. Similarly, MI with stimulus position S_{pos} (Fig. 3C, bottom) is absent before stimulus presentations and saturates to a plateau shortly after stimulus onset. It is higher for the SB than the SU regime, as SB provides a larger dynamic range of responses and sharpened bumps. Stimulus encoding is slightly delayed in SB, since the rearrangement of the bump positions is not as fast as the sudden, stimulus-evoked bump in SU. In the SU regime, all stimulus-related information vanishes shortly after stimulus offset. In contrast to SU, information about stimulus position remains present after stimulus offset in the SB regime, as the stimulus-evoked bump self-sustains itself in its (new) position. It may slowly drift away under the influence of background noise over timescales longer than the observation window considered here.

Extended Data Figure 3-1*B*,*C* show that spatial maps of MI not only depend on time but also on spatial location. Stripes are clearly visible in these "infogram surfaces," because our design comprises only a discrete number of possible stimulus orientations (compare Extended Data Fig. 3-1*A*). Encoding of stimulus features is in general stronger in the bump centers as these locations have a larger dynamic range.

In summary, the normalized MI of firing rate with stimulus provides an interpretable marker of the IPP of "carrying" stimulus (un-)specific information.

IPP analysis can track the loading and maintenance of a representation in working memory

We are able to detect that the poststimulus activity in the SB regime still "carries" stimulus position information. Through which primitive processing operations can this representation be held in working memory? Answering this question requires turning to information dynamics metrics, such as AIS (see Methods), which quantifies the fraction of the information carried by a node's activity at a time t that was already carried at an earlier time t-τ, that is, MI($R(t)$; $R(t$-$\tau)$), with delay $\tau = 40\ \delta t$. Plain AIS tracks the time-lagged maintenance of information by neural activity, irrespective of the origin of this information (stimulus-related or intrinsically given). The process through which this fraction of information is maintained corresponds to the IPP of "buffering" (compare the second cartoon from the top in Figs. 1*B*, 3*D*, left). Note that we focus on buffering by one node only (as we did with "carrying"). Strictly speaking, however, information about stimulus may be carried (redundantly and/or cooperatively) by more than one node simultaneously. Accounting for these more complex scenarios would require using more advanced metrics of storage that discriminate storage of redundant versus synergistic information (see Discussion), which we do not consider here, for the sake of simplicity.

Figure 3*D* shows averaged time traces of AIS computed with the ring model in Figure 3*A–C*. In the SU regime, AIS is positive only during stimulus presentation. It relaxes back to zero after stimulus offset as the stimulus-evoked bumps dissolve back to homogeneous baseline activation. In the baseline pre- and poststimulus presence, all entropy is due to spatially and temporally uncorrelated noise, which is by construction memory-less, thus resulting in null active storage.

The situation is different in the SB regime, in which (spontaneous) bump formation is associated with active storage and thus positive at baseline and after stimulus offset. However, AIS drops at stimulus onset and offset. These events induce changes in activity that cannot be predicted based on prior intrinsic activity and hence convey information, which is not the outcome of "buffering" but must come from outside the system. This information injection is captured by another information-theoretical metric, $\text{TE}_{S \rightarrow R(t)}$ from stimulus to rate (see Methods), tracking the complementary IPP of "transferring." In general, TE identifies the information flow between time series when a given signal Y is influenced by another signal X. It is defined as the conditional MI between the present of Y and the past of X, conditioned on (i.e., factoring out) the past of Y. As shown in Figure 3*E*, the TE peaks match the drops in AIS visible in the middle of Figure 3*D*. At stimulus onset, network nodes modify their algorithmic role, reducing their implication in the IPP of "buffering" and becoming the recipients of information conveyed by the IPP of "transferring." Transfer of information from stimulus to activity occurs also at stimulus offset, when a new information injection indicated by a second peak in TE encodes a release , nd to either produce bump dispersion (in SU) or a decrease in firing rate together with readjustments of the bump shapes (in SB). See also Extended Data Figure 3-1*D* for detailed spatial maps showing the nodes that are most strongly affected by externally injected information at different times.

The information buffered at baseline in the SB regime cannot yet be stimulus-specific as the stimulus has not yet been presented (the positions of spontaneously generated bumps in SB are random). To formalize this intuition, we quantify the fraction of information about the stimulus stored by the network nodes, that is, the stimulus-specific active storage. It is the totally stored information minus the part that does not depend on the presented stimulus orientation (see Methods). The averaged time course of stimulus-specific active storage is shown in the rightmost subpanel of Figure 3*D*. Its trace correctly captures that there is no stimulus-specific information buffering prior to stimulus presentation, while it displays a transient increase after stimulus presentation, for SB also during the poststimulus period. Stimulus-specific active storage thus provides a valid metric to track the active maintenance of information relative to a presented stimulus.

In this single-ring example, the dynamics of "carrying" and "buffering" stimulus-related information are very similar, because all information "carried" by a node exists as a form of "buffering" well after stimulus onset and offset. However, they are not completely identical. In SB, the rebound after stimulus offset in the curve for buffering is the most obvious difference to the carrying curve. It reflects the fact that some of the transient variations in "carrying" are due to "transfer" (compare Fig. 3*E*). In general, compared to the IPP of carrying, buffering is always delayed, since only information already carried by the system can be buffered.

At the functional level, such stimulus-specific maintenance marks the implementation of working memory. At the algorithmic level, our IPP analyses allow a decomposition of working memory: it arises via the loading of stimulus-specific information—through the IPP of "transferring"—into the system's units. By virtue of their collective dynamics, these units are intrinsically devoted to the IPP of "buffering." This algorithmic decomposition provides not only a narrative of how a system's dynamics translates into a function but also yields a quantitative characterization: suitable information-theoretical metrics—AIS for "buffering" and TE for "transferring"—provide a precise evaluation of when, where, and how distinct IPPs are performed.

IPP analysis can track the propagation of representations through a multiregional hierarchy

We now tackle the algorithmic decomposition of the function of activity propagation. As detailed in Models and Methods, we simulated a feed-forward chain of three-ring modules, representing three hierarchically ordered regions (e.g., V1, V2, and above; Fig. 4*A*, bottom). The bottom ring R1 represents a sensory cortical area. It receives an input stimulus, which is sent to hierarchically higher cortical areas (R2 and R3). Each unit in the bottom and middle rings (R1 and R2) is coupled to the corresponding unit (and its local neighbors) in the subsequent rings (R2 and R3), respectively.

Figure 4*B* shows a representative example of single-trial firing rate traces (together with the associated spatial maps of activity) for all three rings, for both the SU (top) and SB regimes (bottom). Red lines and curves indicate units at the position of stimulus injection, blue units far from it, and magenta lines and curves indicate the initial bump position in SB simulations. All panels show the propagation of activity bumps through the hierarchy of rings. Again, bumps are purely stimulus-evoked in the SU regime, while they emerge spontaneously (and are persistent) in the SB regime. The bump's maximum amplitude decreases, and its peak latency is delayed when propagating from the bottom to the top ring. This effect is more pronounced in the SU regime than in the SB regime, where self-amplification via local recurrent excitation acts as a facilitator for propagation. In SB, the effect of forward coupling is already observable without any stimulation: the intrinsic bump positions (magenta lines and curves before stimulus onset) are very similar in the three rings, whereas they would be completely decorrelated if rings were uncoupled.

As for the one-ring model, we study whether bump activity performs the basic IPP of "carrying" information about the stimulus. The encoding dynamics revealed by the MI analyses in Figure 4*C* closely mirror the dynamics of firing rates in Figure 4*B*. The peak amount of information carried about stimulus position is larger in the bottom ring (black curve) and weaker in the top ring (dotted curve). Furthermore, the rise of encoded information is slower and delayed in rings R2 and R3, particularly in the SB regime where the realignment of bump positions is slow and continues in higher-order rings even after stimulus offset. Our model thus successfully captures the propagation of sensory representations.

The obvious IPP algorithmically mediating this propagation is that of "transferring" (compare Fig. 3*E*). In Figure 4*D*, we show the time series of interregional information transfer evaluated via TE (again, see Discussion for some limitations of this metric). TE quickly rises after stimulus presentation, reaching a peak when MI with the stimulus saturates at its maximum plateau value (compare Fig. 4*C*). Transfer is stronger and faster from R1 to R2 than from R2 to R3, again in agreement with the firing rate dynamics in Figure 4*B*. After the peak, transfer drops to a plateau level, which slowly decays after stimulus offset. The profile of transfer is more complex for the SB than for the SU regime. Firstly, in SB, there is inter-ring transfer of information prior to stimulus onset, since the bump positions in R1 and R2 influence those in R2 and R3, respectively (compare

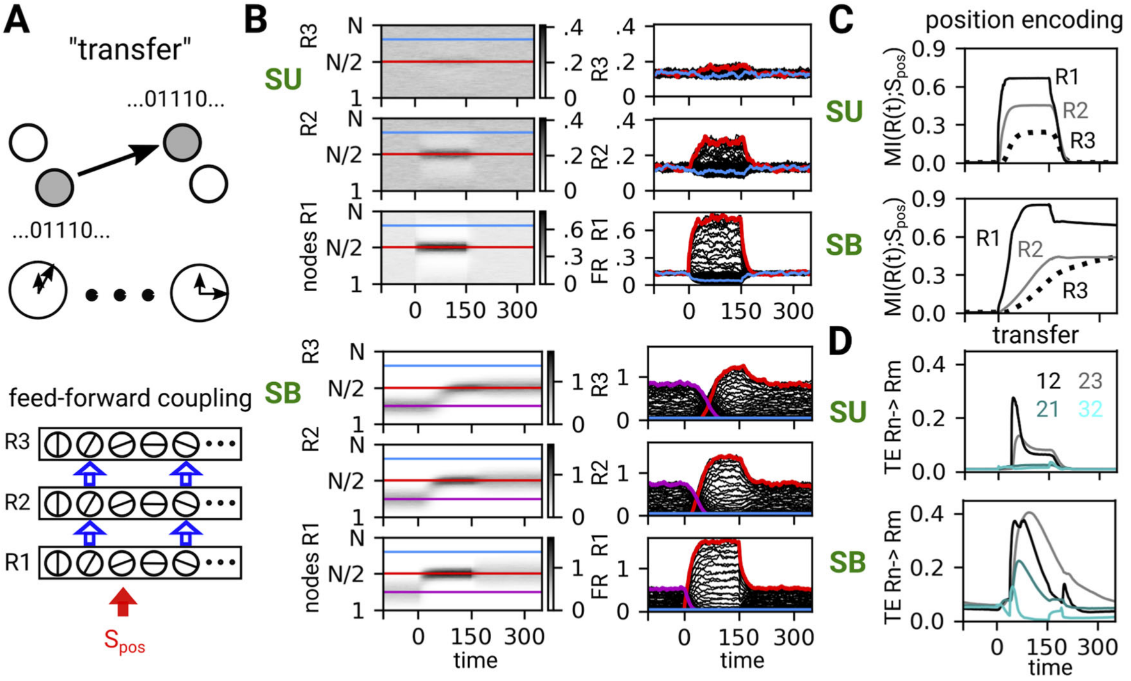

Figure 4. Information transfer in multiregional feed-forward circuits. ***A***, We study the IPP of "transferring" as it gives rise to stimulus propagation across a chain of three feed-forwarded connected regions, each modeled by a different ring network. Only the bottom ring (R1) directly receives stimulus-related inputs (red arrow). ***B***, Spatial maps (left) of single-trial firing rates and corresponding rate time series (right) in R1, R2, and R3 (top, SU regime; bottom, SB regime). Red lines indicate nodes located at the stimulus center, blue lines nodes far from it, and magenta lines indicate the initial bump position for SB, as in Figure 3*B*. ***C***, Time courses of relative MI between rates and stimulus feature (trial and node averages, entropy normalized) reveal stimulus position encoding, transient in SU (top) and persistent in SB (bottom), progressively weaker and more delayed ascending from R1 (black line) to R2 (gray line) and R3 (dotted line). ***D***, Time courses of information transfer, quantified by TE (trial and node averages, entropy normalized) from R1 to R2 (black), R2 to R3 (gray), R2 to R1 (dark cyan), and R3 to R2 (cyan; SU on the top, SB on the bottom). See also Extended Data Figure 4-1, notably for an analysis of the dependency of TE on lag and a comparison with time-lagged MI.

Fig. 4*B*). Secondly, SB curves are broader with marked secondary peaks, associated with the rearrangement of bump positions. Their rearrangement takes longer than their mere creation. The spatial maps of TE in Extended Data Figure 4-1*A* show that substantial transfer occurs even to units far from the stimulus centers as the generation or drift of bumps at locations misaligned with the stimulus must be actively controlled (another interregional functional interaction that TE is able to track).

Since the wiring of the multiregional circuit is purely feed-forward, there should not be any significant feedback information transfer. In the SU regime, the backward TE from higher-order toward lower-order rings is close to zero. However, in the SB regime, a finite backward TE is detected. It is clearly smaller than the forward TE, allowing it to correctly capture the dominant direction of transfer. This spurious backward transfer is due to the misestimation of joint probability density given the finite amount of data, as well as to systematic biases of our simple “plug-in” estimators. Since the wiring of the multiregional circuit is purely feed-forward, there should not be any significant feedback information transfer. In the SU regime, the backward TE from higher-order toward lower-order rings is close to zero. However, in the SB regime, a finite backward TE is detected. It is clearly smaller than the forward TE, allowing it to correctly capture the dominant direction of transfer. This spurious backward transfer is due to the misestimation of joint probability density given the finite amount of data, as well as to systematic biases of our simple “plug-in” estimators. It can be reduced by using a longer delay in estimating TE (Extended Data Fig. 4-1*B*), and it vanishes almost completely when using multivariate delay coordinate embeddings as originally prescribed for TE (Schreiber, 2000; see Extended Data Fig. 4-1*C*). Both approaches allow limiting the impact of fast transients not properly captured by our quantized estimation of joint activity distributions (see Methods). The spurious backward transfer occurs primarily in the particular case of three coupled rings in the SB state: here, the changes in the internal dynamics (bump displacement) are very slow compared to the fast change in the external input (stimulus onset), because the reformation of bumps inside each ring takes a certain amount of time and is even more delayed in rings 2 and 3. In this case, a multi-delay estimator of TE can be used to substantially reduce the bias. In all presented cases, the conclusions reached by more sophisticated (and computationally costly) estimators are qualitatively equivalent to the ones of simpler single-delay TE. Single-delay TE, furthermore, despite its simplicity, already allows a clearer detection of the dominant direction of information propagation than plain time-lagged MI (Extended Data Fig. 4-1*D*), in line with previously published theoretical analyses (Kirst et al., 2016).

In summary, all estimators were able to correctly detect the existence of dominantly feed-forward information transfer. The “transfer” IPP is always the main component in the algorithmic decomposition of the propagation of a sensory representation.

IPP analysis can track the integration of bottom-up and top-down information flows

Our last model configuration is specifically designed to reproduce another important cognitive function: selective attention and the involvement of working memory in its implementation. An attentional effect is depicted as boosting responses to stimuli with attended features and suppressing responses to stimuli with features far from the attended ones (feature–gain–similarity principle, Maunsell and Treue, 2006). Seminal modeling work by Ardid et al. (2007) has first shown that the attentional effects on sensory responses can be explained as a byproduct of the nonlinear integration of sensory bottom-up inputs and top-down inputs from a higher-order region with working memory. This is in line with earlier hypotheses that working memory could be a fundamental component of mechanisms mediating attentional modulation (Desimone and Duncan, 1995). In the following, we will show that this merging of bottom-up and top-down influences can be tracked by the IPP of “integrating,” quantified by synergistic information modification (Fig. 5*A*). In the model proposed by Ardid and co-workers, two ring networks are reciprocally coupled (2RC architecture; see Methods). The lower-order ring R1 represents a sensory area tuned in SU, thus generating stimulus-driven bumps upon stimulus injection. The higher-order ring R2 represents the prefrontal cortex, conditionally set to be in the SU or SB regime, respectively, depending on the attention state “off” (att-OFF) or “on” (att-ON). With att-ON, the second ring is enabled to sustain an induced representation of a presented stimulus, even after the stimulus is removed (i.e., it can act as working memory).

Following Ardid et al. (2007) in the main aspects, we simulate a classic delayed match-to-sample task (Fig. 5*B*). This virtual task mimics actual experiments probing the response of cells recorded in the MT cortex to drifting dot patterns. Cells in MT show a strongly selective response to stimuli drifting in their preferred angular direction (Albright, 1984), resulting typically in bell-shaped tuning curves with a marked unique peak. In our computational model, this selectivity is captured by the heterogeneous responses of units along the sensory ring, resulting in the response profile given by the black curve in the top panel of Figure 5*C*. In the virtual task design of Figure 5*B*, two types of trials exist. The “att-OFF” trials correspond to an empirical condition in which the subject actively attends to a stimulus presented outside the receptive field of the recorded cells [called “unattend” in Treue (2001)]. The “att-ON” trials conversely correspond to the empirical “attended” condition where the attentional spotlight is in the receptive field of the recorded cells. In both conditions, a first cue stimulus is shown and then removed (Fig. 5*B*, red stage), followed by a delay period of a certain length (Fig. 5*B*, black stage) with no stimulus. Then, a second stimulus is presented, whose direction can be close to or far from the direction of the initially cued stimulus (Fig. 5*B*, match stage, magenta). Individual simulated trials for different cue and match stimulus configurations are shown in Figure 5*D*, in both the att-ON (top) and att-OFF (bottom) conditions.

When simulating neural responses in the att-OFF condition (prefrontal area ring R2 tuned to SU), the response profile of the sensory ring to the match stimulus is unchanged with respect to the one during the cue stage. The presentation of match stimuli with different directions simply produces rotated response profiles (Fig. 5*C*, black and dashed gray lines). Inspecting the responses in individual trials, we see that all match stage activity bumps in the sensory ring look similar and that the activity in the prefrontal ring R2 has a smaller rate and a worse signal-to-noise ratio compared to R1 (Fig. 5*D*, bottom).

In contrast, in the att-ON condition (a copy of the cue has been held through the delay period by R2 switched to SB), the response profile at match stimulus is differently modulated depending on the relative difference of orientation between cue and match stimuli (Fig. 5*C*, $\Delta\Theta$). If cue and match stimuli have identical orientations ($\Delta\Theta = 0$), the response of the direction-selective units in R1 is boosted, while the response of units selective to stimuli far from the attended one is reduced. This can be seen in single-trial responses (Fig. 5*D*, top), where the match

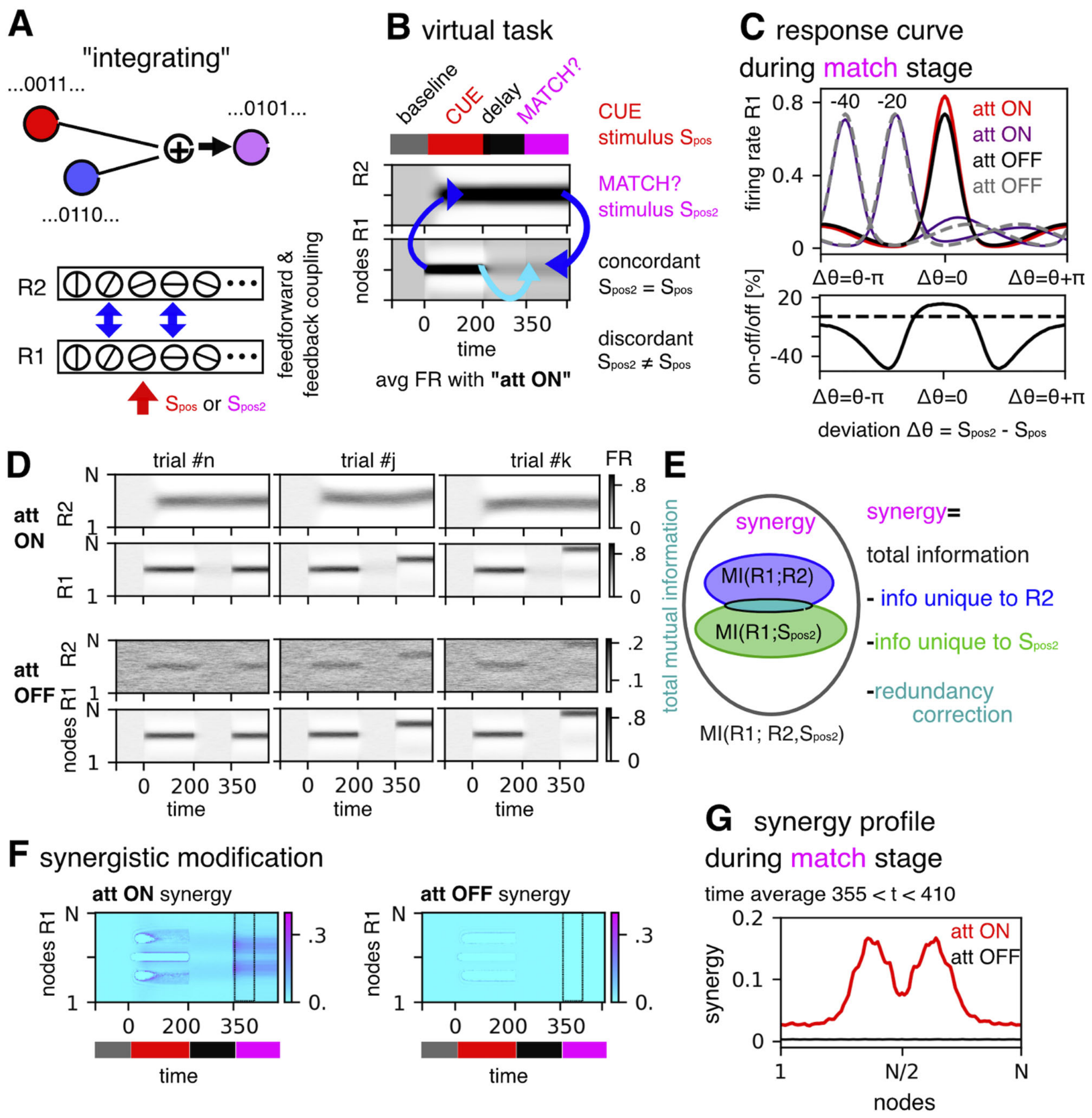

Figure 5. Information integration and synergy in the presence of top-down attentional modulation. ***A***, We study the IPP of "integrating" as it mediates the emergence of top-down attention-like modulation of stimulus response in a bi-regional circuit, composed of two reciprocally coupled ring networks, representing respectively a low hierarchical order sensory region and a higher-order frontal region. ***B***, In the virtual task we simulate, after a baseline period, two stimulus presentations, during a cue and match stage, respectively, with positions S_{pos} and S_{pos2} (red arrow), separated by a delay period without stimulation. Such a configuration mimics a selective attention experiment in which a copy of the presented stimulus is uploaded to a frontal working memory module (upward blue arrow), which stores it actively through the extent of the delay period (light blue arrow). At the moment of the match, this working memory copy interacts (downward blue arrow) with the sensory representation evoked by the newly presented stimulus, matching or not the previously cued direction. The circuit can be set into an "attention-ON" (upper ring in SB regime) or an "attention-OFF" (upper ring in SU mode) condition. Panel ***B*** shows trial-averaged spatial maps in the att-ON condition for matching stimuli directions ($S_{pos} = S_{pos2}$). ***C***, Response curves of firing rates averaged over trials and time during S_{pos2} presentation (match stage). The red curve for match trials ($S_{pos} = S_{pos2}$), purple curves for no-match, and both for att-ON. In att-OFF, the black curve corresponds to a match, otherwise dotted gray curves. Bottom: attentional modulation index showing the percent enhancement (or depression) of firing rate during match stage in att-ON versus att-OFF conditions. ***D***, Firing rate spatial maps for three single trials with different configurations of S_{pos} and S_{pos2}), in bottom (R1) and top ring (R2; top, att-ON; bottom, att-OFF conditions). ***E***, Venn diagram indicating the PID of the total MI between the sensory response in R1 and the pair of bottom-up sensory and top-down frontal inputs: synergy equates the fraction of this total which is neither uniquely carried by R2 and S_{pos2}, nor redundant between them. *For details on individual terms of the PID, see Extended Data Figure 5-1.* This synergistic information is extracted by nodes in R1 through the process of information modification. ***F***, Spatial map of the synergistic modification (normalized) in ring R1 in attention-ON (left) and OFF (right) conditions. Synergy is much stronger in att-ON condition, particularly in the match stage. ***G***, Section of the synergy surface during the early match stage (section averaged over the time window indicated in ***F***, by the dotted black rectangle).

bump is slightly darker for identical cue and match configuration (trial #*n*). It is clearer in the red activation profile in Figure 5*C* whose peak at $\Delta\Theta = 0$ is higher than the black one obtained for att-OFF. If cue and match stimuli have different directions, the difference in $\Delta\theta$ modulates the corresponding response profile (Fig. 5*C*, purple curves) in the att-ON condition. The net amount of (simulated) attention-induced modulation can be quantified by computing the percent difference ratio between the response profiles to a stimulus in att-OFF and ON conditions as shown in the bottom panel of Figure 5*C*. In our virtual task, the positive modulation can be as large as +15% for responses to identical cue and match stimuli and down to between −10% and even −45% for stimuli, which are ~45° displaced. We now study the algorithmic effects of these nonlinear dynamics.

We focus specifically on the IPP of "integrating," occurring here at the match stage when the sensory response (Fig. 5*D*, R_1) is the byproduct of nonlinearly merged bottom-up stimulus-related input (S_{pos2}) and top-down attention-related input (R_2). The information-theoretical quantity, we use to track this IPP is synergistic information modification (Lizier et al., 2013). As graphically depicted in the information Venn diagram of Figure 5*E*, the two bottom-up S_{pos2} and top-down R_2 inputs carry together (when considered jointly) a certain amount of information $\text{MI}_{\text{tot}} = \text{MI}((R_2, S_{\text{pos2}}); R_1)$ about what is going to be the output sensory response R_1. A fraction of this total information is contributed uniquely by each of the two considered inputs. At match stimulus onset, only the bottom-up input S_{pos2} can convey information about the direction of the newly shown match stimulus, while only the top-down input R_2 can carry information about the previously presented cue stimulus. Both inputs contribute to determining the final output response, and they comprise two unique information fractions conveyed exclusively by each of the two inputs (Fig. 5*E*, green and blue areas). Some additional information may be shared between the two inputs—including noise entropy—as captured by the redundant information fraction (Fig. 5*E*, cyan intersection). Yet, the sum of unique and redundant contributions could be smaller than the total information MI_{tot}. Some information necessary to determine the response could be conveyed by the two inputs in combination but by neither of them in isolation. This surplus contribution —"more than the sum of the parts" (Anderson, 1972)—is the synergistic fraction of the total information (Fig. 5*E*, white area), and its extraction by the output nodes is termed the synergistic modification operation. We estimate these unique, redundant, and synergistic contributions, quantifying when and where along the virtual task (Fig. 5*B*) the activity of the interacting rings implements the IPP of "integrating."

This extraction of the synergistic information is tracked and quantified by the information modification surfaces shown in Figure 5*F*, for both att-ON (left) and att-OFF (right) conditions (compare Extended Data Fig. 5-1 for details on the individual terms contributing to its computation). In addition, Figure 5*G* shows the profile of a section of the modification surface at the beginning of the match stage (Fig. 5*F*, averaging range delimited by the dotted black rectangle). Clearly, sensory ring units perform information modification at specific task-related locations and times in the att-ON condition while there is nearly no synergistic modification for att-OFF.

The most prominent involvement in information modification occurs in the match stage, particularly at the immediate onset of the match. This is precisely when attentional modulations of stimulus response occur. As visible when comparing the profiles of attentional modulation (Fig. 5*C*, bottom) and information modification (Fig. 5*G*), the participation of a node in information modification is stronger for the prominent bump in attentional modulations. Specifically, modification is enhanced at the bump flanks, where the most attentional depression is observed. Modification at the bump center position is weaker, because here, the strong net drive helps the recurrent excitatory connectivity within the sensory ring to sustain the boosted activity.

For att-ON, information modification can be seen to occur during different virtual task stages. Modification stripes preceding the match stage occur during the delay stage with a much weaker intensity. They are related to the fact that some small-intensity activity is top-down transferred from the bump in R2 (compare Fig. 5*F*). These nonlinear interactions between the working memory bump during the delay and its "sensory shadow" effectively reduce the variability of the sensory ring activity in att-ON versus att-OFF. This is another type of nonlinear phenomenon beyond rate modulations that can result in modification (see Discussion). Other modification events occur during the cue stage, probably due to the transient reshaping dynamics of activity bumps caused by the parameter changes for R2 to switch from SU to SB regime.

In the att-OFF state, information modification is close to zero and possibly estimated to small positive values, because of numeric estimation artifacts. As detailed by Extended Data Figure 5-1, the surfaces shown in Figure 5*C* are the sum of several other surfaces corresponding to the different terms in the expression for the synergistic information part (see Methods). Numerical errors could thus be more important, as more steps are involved.

In conclusion, the function of selective attention admits an algorithmic decomposition involving the IPP of "integrating," a much more complex function than the IPPs of "carrying," "buffering," or "transferring" described in previous sections.

Information dynamics in a large-scale model of the cerebral cortex

The ring models we have studied so far represent generic coupled brain regions with only stylized notions of hierarchical connectivity (Fig. 4, feed-forward propagation; Fig. 5, top-down integration). Detailed information about interregional cortical connectivity is, however, available and large-scale models of cortical activity have been constructed, embedding empirically derived multiregional connectomes (Deco et al., 2011). In these models, the local activity of individual regions is modeled with simple neural mass equations. Each region receives inputs from other regions, scaled by the relative strength of connections within a connectome matrix. In addition, the distinct fiber tract lengths of the connections can be represented by different propagation delays. To validate our IPP approach beyond generic and ad hoc constrained ring model architectures, we now consider the case of stimulus-evoked activity in a large-scale model of macaque monkey cortex with realistic connectivity (Fig. 6).

Specifically, we re-implemented a model previously published by Joglekar et al. (2018), embedding a directed and weighted connectome derived from systematic tracer experiments (Markov et al., 2014; Fig. 6*A*). The "bow-tie" architecture of such a connectome introduces a barrier to the free propagation of externally injected stimuli. As shown in Figure 6*B*, the application of a brief transient input to region V1, mimicking the presentation of an external visual stimulus, leads to a very limited propagation of activity along the visual stream. However, a GBA mechanism can be introduced, in which lr excitatory connectivity is

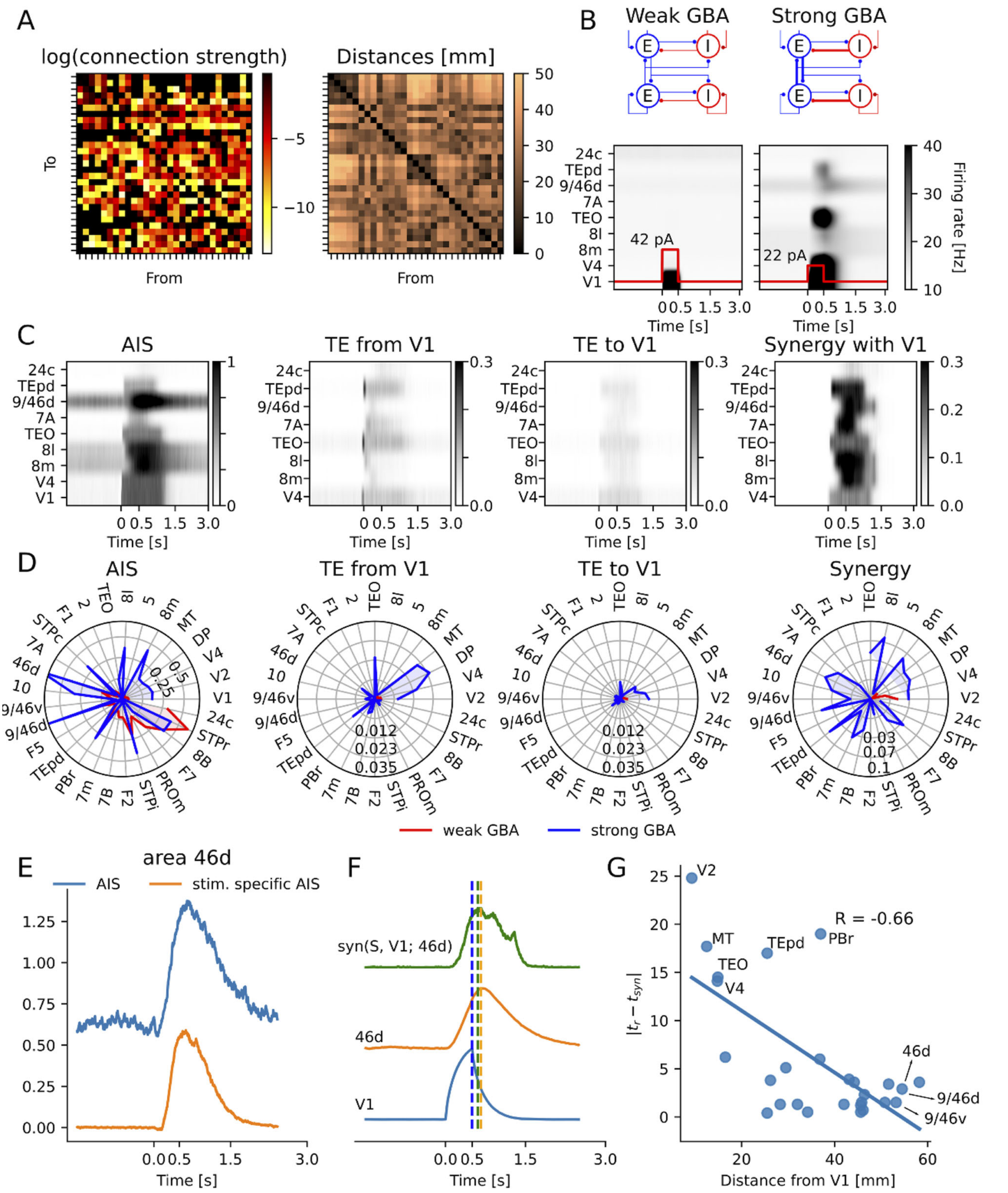

Figure 6. Information dynamics in a large-scale model of the cerebral cortex. Large-scale connectome-based mean-field model of nonhuman primate cerebral cortex. ***A***, Interregional connectivity is described by empirically derived matrices of connection strength (left) and connection length (right), determining efficacy and delay of inputs from remote regions. ***B***, We consider two regimes of the model, a weak and a strong GBA mode. To favor signal propagation, in strong GBA, long-range excitation is strengthened together with an increase of local inhibition to excitation, to avoid rate explosion. Shown at the bottom are rasters of the rate response of different regions of interest in response to a stimulus pulse presented to region V1, in the weak and strong GBA modes. ***C***, Rasters of IPP metrics dynamics associated with the presentation of a stimulus in the strong GBA regime (as in panel ***B***, right). ***D***, Alternative representation of IPP metrics in the form of radar plots (blue, strong GBA; red, weak GBA). From left to right: AIS (stimulus-unspecific, see panel ***E*** for an example of stimulus-specific, in orange, vs stimulus-unspecific, in blue, storage); TE from and to region V1; synergistic integration by region V1 of stimulus-related and top-down inputs from different regions. ***F***, From bottom to top: time series of rate responses of a sensory and a frontal region, together with the synergistic integration by V1 of the bottom-up stimulus and top-down frontal inputs. Peak positions are highlighted by dashed lines. ***G***, Scatter plot of synergistic modification peak latency versus distance of top-down input source from V1 reveals an anticorrelation (significant Pearson's *R*, 95% confidence intervals −0.658 and −0.643).

enhanced, while simultaneously strengthening local inhibition (to excitatory populations within each region; Fig. 6*B*, top right). In this way, propagation is facilitated by stronger connections, while the increased amount of excitation is controlled by a modified inhibitory-to-excitatory balance. Figure 6*B* compares a scenario of weak (left) versus strong GBA (right), revealing that stimulus-related activity volleys can reach higher-order regions only in the latter case. Following Joglekar et al. (2018), higher-order regions are endowed with stronger excitability and correspondingly longer time constants. Thus, once a stimulus-related volley has reached a higher-order region (e.g., region 46d), the stimulus-evoked activity will last well beyond the offset of the applied stimulus. Adoption of a strong GBA regime and hierarchy-scaled excitabilities allows thus to implement in the large-scale model a scenario reminiscent of the att-ON condition in Figure 5, where the top ring, after being cued to attend a stimulus direction, is switched in the SB regime, sustaining bumps in absence of a stimulus.

We performed simulations of stimulus propagation in this large-scale cortical model and extracted its algorithmic decomposition, revealing "buffering," "transferring," and "integrating" IPPs. Figure 6*C* shows raster plots of IPP metrics, computed in the strong GBA regime, whereas Figure 6*D* shows an alternative representation of these IPPs, time-averaged radar charts, where blue lines indicate strong and red lines weak GBA regimes. The results of analyzing (stimulus-unspecific) AIS are reported in the first column on the left. We observe that sensory regions (e.g., V1 and V4) and (pre-) frontal regions (e.g., 46d, 8) have very different storage dynamics. The raster plot (Fig. 6*C*) shows that in sensory regions, AIS drops relatively quickly after stimulus offset, while in higher-order regions, it persists for a longer time span, an effect of the longer intrinsic time constants. Areas 46d and 8 display positive AIS before stimulus presentation, reminiscent of ring networks in the SB regime (compare Fig. 3*B*). However, when additionally computing stimulus-specific storage (Fig. 4*E*, area 46d), we see that the fraction of storage imputable to stimulus is zero before stimulus onset and increases thereafter. In the low GBA condition, only a limited set of regions displays noticeable AIS (Fig. 4*D*).

The IPP of transfer is shown in the second and third columns of Figure 6, *C* and *D*. Here, we consider the forward TE from V1 to other areas and the backward transfer to V1. In the strong GBA regime, V1 transfers information to a variety of other regions, mostly in the ventral (e.g., V4) or dorsal (e.g., MT) visual stream. Backward transfer is weaker and arises mostly from nearby low-level visual regions (e.g., V2 and V4). Again, a transfer is considerably disabled in the low GBA regime (Fig. 4*D*).

Finally, we consider V1 as a target node integrating the bottom-up stimulus-related input and top-down inputs from other regions in the large-scale cortical circuit and compute the amount of synergistic modification performed by V1 (Fig. 6*C*,*D*, rightmost columns). Strong stimulus-related synergies exist with a variety of regions through the entire cortical hierarchy, particularly strong with frontal and prefrontal regions (areas 8 and 46d). Similar to the IPPs of "buffering" and "transferring," a strong GBA regime is needed to establish substantial synergistic modification. In Figure 6*F*, we study in detail the latencies with which synergistic integration arises. Figure 6*F* (left) shows the synergistic information in the stimulus-induced V1 response due to the integration of top-down inputs from region 46d. As expected, the V1 response peak precedes the peak in the 46d response, which lasts longer than the V1 response. Synergy starts growing in parallel with the increase of activity in the area 46d region, and the synergy peak occurs very close to the area 46d response peak, indicating a fast integration of top-down signals by V1. Remarkably, as shown by the scatter plot in Figure 6*G*, synergistic integration is achieved first with top-down inputs from regions hierarchically distant from V1. Indeed, a significant anticorrelation ($R = -0.66$, $p = 0.0002$) exists between distances in the matrix of Figure 6*A* (right) and the peak of the synergy time course. This finding is nontrivial and reveals how the bow-tie topology of the corticocortical connectome supports a fast-"forward" transfer of stimulus-related information followed by a "backward" integration of top-down influences, initiated by the top-most regions (see Discussion).

Discussion

Information processing in cognitive sciences is commonly viewed in terms of box–arrow models linking perception to behavior, not necessarily with explicit reference to neural mechanisms (Fodor, 1968; Rumelhart and McClelland, 1986) but increasingly so (McClelland and Lambón Ralph, 2013), also due to the rise of neuroimaging (Price, 2018). Hypotheses about processing architectures are validated through experimental tasks designed to disentangle the relative contributions of different boxes in such models. It is difficult, however, to interpret the results without implicit reference to concepts of the postulated theory (Cooper, 2007). If this theory deviates strongly from the (unknown) underlying neuronal computations, the resulting analyses may be inherently biased. There is thus a need for data-driven, agnostic approaches to get access to the algorithmic level.

We propose a set of metrics to detect and measure elementary processing operations when applied to the analysis of neural activity. The rigor in the definition of these IPPs (buffering, transferring, and integrating) necessitates that they must be abstract and act in plain, identifiable ways on information conveyed by neural activity. Although these operations are far from evident cognitive functions (e.g., attentional modulation), they constitute their necessary low-level ingredients, a sort of "neural assembly language." Varying combinations of IPPs build up into a variety of functions, like the instruction set of a conventional digital computer (Wilkes et al., 1951): despite their deceptive simplicity, low-level instructions are sufficient to generate a variety of software outputs, from the word processors we used to write this article to the media players that have distracted us during its preparation.

Low-level information processing naturally emerges from the collective dynamics of complex nonlinear systems (Crutchfield & Mitchell, 1995; Packard, 1988; Shaw, 1984). For instance, in cellular automata, like Conway's "Game of Life," dynamical patterns known as "gliders" act as agents of information transfer and their collisions as information modification events (Lizier, 2013). In our study, the IPP of "transferring" is materialized by volleys of propagating activity in coupled ring networks (Fig. 4)—like gliders in the Game of Life—and the IPP of "integrating" by activity volleys colliding within the sensory ring (Fig. 5). In contrast to abstract toy systems, coupled ring models correspond to actual neural circuits mimicking cognitive functions. Within this framework, we can thus make a first step toward bridging the gap between Marr's first and third levels, that is, from the structure of the neural circuit to its function. The missing link (Marr's second level) is provided by the algorithmic decomposition of the simulated activity, which precisely quantifies how (through which primitive operations), when (in which epochs during the task), and where (by which network nodes) information is processed. In this respect, our study is based on a ground-truth model: knowing both the circuit wiring and the emulated cognitive function with precision and certainty allows us to determine the cocktail of IPPs associated with a certain function. The measured spatiotemporal patterns of IPP recruitment are compatible with reasonable, a priori expectations, confirming that the output results of our analyses are trustworthy.

Next, we apply our metrics to the nonhuman primate connectome model (Fig. 6), which represents many cortical regions with realistic connectivity. The structural connectivity for this model is characterized by a double-hierarchical organization despite its heterogeneous and irregular structure. Those characteristics allow one to formulate hypotheses on how computations are performed only up to some extent since they are not prescribed by an imposed custom architecture.

As shown by Joglekar et al. (2018), the forward hierarchy can serve as a substrate for forward propagation of stimulus-related information to high-order regions, which is here tracked by TE. After reaching the frontal subnetworks, their longer time constants allow the reverberation of this information in working memory, as tracked by AIS. Our IPP analyses also allow identifying another algorithmic effect of the bow-tie connectome, now serving as a substrate for the synergistic integration of top-down inputs. Similar to the two reciprocally coupled rings, top-down modulations of V1 activity are an automated consequence of reverberant stimulus-related activity in high-order regions. Thus, analyzing IPPs provides a picture of the algorithmic effects of stimulus propagation in a realistic connectome model. It even allows disentangling the detailed timing of three distinct primitive computations of “transferring,” “buffering,” and “integrating,” in relation to the ascending and descending hierarchy within the connectome.

The anticorrelation between hierarchical distance and synergy maximum revealed in Figure 6*F* is an intriguing prediction, supporting the role of frontal and prefrontal regions as leading top-down controllers (Paneri and Gregoriou, 2017). Although these regions are “far away” from V1, they are the first to exert a measurable information modification effect in V1. Additional information modification in V1 then occurs via the integration of inputs from other “closer” cortical regions, however, only with greater latency, as the activity of these regions must also be first modified as an influence of top-down influences from the top of the hierarchy. Such anticorrelation is thus interpretable and reasonable. It was, however, unanticipated and makes us optimistic about the heuristic potential that IPP analyses may have when applied agnostically to large-scale activity recordings.

Our proposal to seek IPPs underlying cognitive computations is not completely novel (Lungarella and Sporns, 2006; Ince et al., 2015). Training in specific tasks has been shown to automatically confer superior performance in apparently unrelated tasks (Singley and Anderson, 1989). This finding led to the speculation that cognitive algorithms may involve shared processing subroutines so that training of low-level processes explains the transfer of cognitive skills across tasks. Such a notion of “primitive elements” of cognitive processing (Taatgen, 2013) is naturally algorithmic, since it refers to “prefunctional” information manipulations, participating in the *implementation* of the final function. Analogously, other cognitive theories postulate the existence of intermediate representations (Wickelgren, 1999; Mel and Fiser, 2000) between the encoding of isolated parts of objects (e.g., contour segments) and the fully integrated encoding of whole objects (e.g., shapes). Such intermediate representations could be reinterpreted as primitive algorithmic steps toward object recognition. Our notion of IPPs lies at an even lower hierarchical level, since such prefunctional cognitive operations could themselves be further decomposed into IPPs.

Another asset of our IPP analysis is that it goes beyond the conventional study of functional connectivity. The latter just detects which units process information together, while IPPs additionally reveal the qualitative type of processing. For instance, very similar functional connectivity motifs (activity bumps in the sensory ring) are generated in attend-ON and attend-OFF conditions during a match in the simulated experiment of Figure 5. Functional connectivity primarily captures information *transfer* for both conditions, while IPPs additionally detect substantial information *modification* for attend-ON, revealing two *qualitatively* distinct modes of processing. The IPP framework thus enables stronger constraints on hypotheses about cognitive processing implementations. The capacity to simultaneously track different types of processing across different locations will facilitate the identification of putative cognitive architectures combining parallel and sequential aspects (Zylberberg et al., 2011). Information decomposition techniques can also be used to reveal distinct information processing roles in different cognitive domains (Luppi et al., 2022).

The rate models considered here are extremely simplified with respect to biological neural circuits. They simply serve to generate activity time series from generic neural systems mimicking cognitive functions, without bothering too much about realism. Despite their simplicity, ring models can generate a huge variety of dynamics. We focus on asynchronous activity regimes, however, enhancing inhibition or varying delays gives rise to alternative regimes characterized by oscillatory activity, including traveling waves or even chaotic oscillations (Roxin et al., 2005, 2006; Ardid et al., 2010; Battaglia and Hansel, 2011). In perspective, oscillatory regimes could be used to quantify whether their presence affects primitive computations (e.g., boosting modification and/or transfer), to elucidate the effects of cortical traveling waves (Gong and van Leeuwen, 2009; Muller et al., 2018; Chemla et al., 2019) on information processing (colliding wavefronts might act as information modification events), or to benchmark tools for spectrally resolved information-theoretical analyses (Pinzuti et al., 2020).

Like previous studies of state-dependent information transfer (Battaglia et al., 2012; Palmigiano et al., 2017), we capitalize on the possibility of simulating arbitrarily large quantities of data in perfectly controlled conditions to enable a straightforward (binning) estimation of information-theoretical functionals. Still, estimation is error-prone, as revealed by the spurious inference of information transfer from higher- to lower-order rings in the feed-forward configuration of Figure 4. The correct qualitative conclusion is achieved (i.e., the dominant direction of transfer), but these results give a warning about the use of estimators. Binning strongly depends on the number of samples, is biased, and suffers from the curse of dimensionality (Treves and Panzeri, 1995; Panzeri et al., 2007). For the analyses shown in Extended Data Figure 4-1*B*,*C*, we used an alternative based on semi-parametric estimation techniques, namely, the GCMI (Ince et al., 2017).

Note that we use a mixture of classical metrics and the modern PID approach. Both can have certain deficiencies, especially with respect to their interpretation. For example, our “classic” definitions of active storage and TE may conflate synergetic and unique information [see Barrett (2015) for the case of Gaussian systems; compare also Gutknecht et al. (2021)]. TE may, under certain circumstances, overestimate the information flow (James et al., 2016). Another concern could be caused by the pairwise analysis, rather than multivariate, to calculate transfer. However, in a system of many components, a transfer may occur groupwise [“polyadic interactions” by James et al. (2016)]. Yet, the network approach provides a useful first approximation to track dynamics (Bullmore and Sporns, 2009; Kirst et al., 2016), as long as we are aware that higher-order interactions may be present (Davison et al., 2015). Yet another limitation of our approach is the use of the MMI implementation of PID that may overestimate the amount of redundancy or synergy, given that it just provides upper bounds for them. For instance, the redundancy between two Gaussian inputs both carrying some information on a third univariate output can be proved to be positive in the MMI ansatz, even if the two input variables are completely uncorrelated (Barrett, 2015), and this redundancy misestimation necessarily leads to synergy misestimation. Superior metrics have been proposed; see Bertschinger et al. (2014) for alternatives to the MMI ansatz in special cases and James et al. (2018) and Kay and Ince (2018) for formulations of PID in terms of dependency constraints decomposition (available for both discrete and Gaussian variables). Eventually, even superior evaluation of unique information beyond MMI and explicit accounting of synergies never fully protect against the possibility that paradoxical cases arise. The extreme scenarios introduced in the theoretical literature to probe the limits of different paradigms may be unlikely to occur in actual neural activity recordings (and simulations from realistic models). Nevertheless, in both “legacy” and “modern” approaches, the interpretations of information flow, transfer, synergy, redundant and unique information, etc. should always be taken with due caution.

Here, we do not address open questions regarding the most adequate estimators to compute IPPs, which is beyond the scope of this study. We propose instead a pragmatic set of measures that helps extract information about neural computing from data. Nevertheless, for the specific case of our results, constructed from real mechanistic circuit simulations rather than abstract statistical models (conceived to test metric limits), the estimated synergistic modification maps are perfectly interpretable, making us more confident that we are tracking genuinely nontrivial computations. Our methodological choices have at least the advantage of being widely used [compare the use of the MMI ansatz by Luppi et al. (2022)] and simple to implement.

The set of IPPs we propose is far from exhaustive. PID exists for more than three variables, yielding a combinatorially exploding number of information “atoms” (Williams and Beer, 2010). Multivariate frameworks to quantify the informational effects of emergent collective behavior have been proposed for arbitrarily large systems (Rosas et al., 2020). Any additional IPP would still capture the informational effects of coordinated dynamics. The name “dynome” has been proposed for the collection of possible dynamic modes that a neural circuit with a given connectome can support (Kopell et al., 2014). In our algorithmic-centered view, every dynamical mode within the “dynome” could map to an element within the “infome,” a repertoire of operators on information conveyed by the corresponding dynamics. Possible examples are transiently coherent oscillatory bursts, serving as information routing enablers (Palmigiano et al., 2017), or the recruitment of distinct subsystems with identical neurons processing information differently at different times (Clawson et al., 2019; Pedreschi et al., 2020). In general, the number of considered IPPs should be tailored to match the variety of intrinsic activity patterns that the considered neural circuit engenders.

Until now, most attempts to identify canonical computations, for example, inhibition-driven rerouting or normalization (Pouille and Scanziani, 2004; Carandini and Heeger, 2011; Hangya et al., 2014; Miller, 2016), were committed to specific connectivity motifs being responsible for specific processing types. Such structure-centric views may be limited by the fact that a connectivity motif can behave differently in different contexts (Aertsen et al., 1989; Nadim et al., 2008, Kirst et al., 2016), so that a single motif could perform multiple computations. Alternatively, it is possible that different connectivities implement similar dynamics (Marder and Goaillard, 2006, Yger et al., 2011, Voges and Perrinet, 2012), resulting in similar computations. The quantification of IPPs allows the study of information processing directly at the algorithmic level, in a way commensurable with, but “disembodied” from the specific circuit mechanisms producing it. This may allow the detection of specific cognitive processes (e.g., attentional modulation) through the identification of their informational signatures (e.g., boosted information modification). IPP analyses may, also in the absence of (apparent) damage in the underlying circuits, allow the detection of disruptions in the information processing itself, thus providing a fundamental “software” explanation for cognitive impairments in pathologies (Clawson et al., 2023).

References

Aertsen AM, Gerstein GL, Habib MK, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”. J Neurophysiol 61:900–917.

Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 52:1106–1130.

Anderson PW (1972) More is different. Science 177:393–396.

Ardid S, Wang X-J, Compte A (2007) An integrated microcircuit model of attentional processing in the neocortex. J Neurosci 27:8486–8495.

Ardid S, Wang X-J, Gomez-Cabrero D, Compte A (2010) Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention: gamma-range synchronization between sensory and executive cortical areas. J Neurosci 30:2856–2870.

Barrett AB (2015) Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. Phys Rev E 91:052802.

Battaglia D, Hansel D (2011) Synchronous chaos and broad band gamma rhythm in a minimal multi-layer model of primary visual cortex. PLoS Comp Bio 7:e1002176.

Battaglia D, Witt A, Wolf F, Geisel T (2012) Dynamic effective connectivity of inter-areal brain circuits. PLoS Comput Biol 8:e1002438.

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc Natl Acad Sci U S A 92:3844–3848.

Bertschinger N, Rauh J, Olbrich E, Jost J, Ay N (2014) Quantifying unique information. Entropy 16:2161–2183.

Bím J, De Feo V, Chicharro D, Bieler M, Hanganu-Opatz IL, Brovelli A, Panzeri S (2020) A non-negative measure of feature-related information transfer between neural signals. bioRxiv 758128. https://www.biorxiv.org/content/10.1101/758128v2

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. Neural Comput 12:1531–1552.

Bressler SL, Seth AK (2011) Wiener-Granger causality: a well established methodology. NeuroImage 58:323–329.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. Proc Natl Acad Sci U S A 101: 9849–9854.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62.

Chemla S, Reynaud A, di Volo M, Zerlaut Y, Perrinet L, Destexhe A, Chavane F (2019) Suppressive traveling waves shape representations of illusory motion in primary visual cortex of awake primate. J Neurosci 39:4282–4298.

Clawson W, Vicente AF, Ferraris M, Bernard C, Battaglia D, Quilichini PP (2019) Computing hubs in the hippocampus and cortex. Sci Adv 5: eaax4843.

Clawson W, Waked B, Madec T, Ghestem A, Quilichini PP, Battaglia D, Bernard C (2023) Perturbed information processing complexity in experimental epilepsy. J Neurosci 43:6573–6587.

Combrisson E, Allegra M, Basanisi R, Ince RA, Giordano BL, Bastin J, Brovelli A (2022a) Group-level inference of information-based measures for the analyses of cognitive brain networks from neurophysiological data. NeuroImage 258:119347.

Combrisson E, Basanisi R, Cordeiro VL, Ince RA, Brovelli A (2022b) Frites: a Python package for functional connectivity analysis and group-level statistics of neurophysiological data. J Open Source Softw 27:3842.

Cooper RP (2007) The role of falsification in the development of cognitive architectures: insights from a Lakatosian analysis. Cogn Sci 31:509–533.

Cover T, Thomas J (2006) *Elements of information theory*. Hoboken, NJ: Wiley-Interscience.

Crutchfield JP, Mitchell M (1995) The evolution of emergent computation. Proc Natl Acad Sci U S A 92:10742–10746.

Davison EN, Schlesinger KJ, Bassett DS, Lynall M-E, Miller MB, Grafton ST, Carlson JM (2015) Brain network adaptability across task states. Plos Comput Biol 11:e1004029.

Deco G, Jirsa VK, McIntosh AR (2011) Emerging concepts for the dynamical organization of resting-state activity in the brain. Nat Rev Neurosci 12: 43–56.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

Eppler JM, Helias M, Muller E, Diesmann M, Gewaltig M-O (2008) PyNEST: a convenient interface to the NEST simulator. Front Neuroinform 2:12.

Fodor JA (1968) The appeal to tacit knowledge in psychological explanation. J Philos 65:627–640.

Gong P, van Leeuwen C (2009) Distributed dynamical computation in neural circuits with propagating coherent activity patterns. PLoS Comput Biol 5: e1000611.

Granger C (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438.

Grossman S, Yeagle EM, Harel M, Espinal E, Harpaz R, Noy N, Megevand P, Groppe DM, Mehta AD, Malach R (2019) The noisy brain: power of resting-state fluctuations predicts individual recognition performance. Cell Rep 29:3775–3784.e4.

Gutknecht AJ, Wibral M, Makkeh A (2021) Bits and pieces: understanding information decomposition from part-whole relationships and formal logic. Proc Math Phys Eng Sci 477:20210110.

Hangya B, Pi HJ, Kvitsiani D, Ranade SP, Kepecs A (2014) From circuit motifs to computations: mapping the behavioral repertoire of cortical interneurons. Curr Opin Neurobiol 26:117–124.

Helmstaedter M, Briggman KL, Turaga SC, Jain V, Seung HS, Denk W (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500:168–174.

Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. Hum Brain Mapp 38: 1541–1573.

Ince RAA, Van Rijsbergen NJ, Thut G, Rousselet GA, Gross J, Panzeri S, Schyns PG (2015) Tracing the flow of perceptual features in an algorithmic brain network. Sci Rep 5:17681–17717.

James RG, Barnett N, Crutchfield JP (2016) Information flows? A critique of transfer entropies. Phys Rev Lett 116:238701.

James RG, Emenheiser J, Crutchfield JP (2018) Unique information via dependency constraints. J Phys A: Math Theor 52:014002.

Joglekar MR, Mejias JF, Yang GR, Wang XJ (2018) Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. Neuron 98:222–234.

Kay JW, Ince RAA (2018) Exact partial information decompositions for Gaussian systems based on dependency constraints. Entropy 20:240.

Kirst C, Timme M, Battaglia D (2016) Dynamic information routing in complex networks. Nat Comms 7:11061.

Kopell NJ, Gritton HJ, Whittington MA, Kramer MA (2014) Beyond the connectome: the dynome. Neuron 83:1319–1328.

Latham PE, Nirenberg S (2005) Synergy, redundancy, and independence in population codes, revisited. J Neurosci 25:5195–5206.

Lizier JT (2013) *The local information dynamics of distributed computation in complex systems*. Berlin, Heidelberg: Springer-Verlag.

Lizier JT, Bertschinger N, Jost J, Wibral M (2018) Information decomposition of target effects from multi-source interactions: perspectives on previous, current and future work. Entropy 20:307.

Lizier JT, Flecker B, Williams PL (2013) Towards a synergy-based approach to measuring information modification. In Proc. IEEE Symposium on Artificial Life, 43–51.

Lizier JT, Prokopenko M, Zomaya AY (2012) Local measures of information storage in complex distributed computation. Inf Sci 208:39–54.

Lungarella M, Sporns O (2006) Mapping information flow in sensorimotor networks. PLoS Comp Biol 2:e144.

Luppi AI, Mediano PA, Rosas FE, Holland N, Fryer TD, O'Brien JT, Rowe JB, Menon DK, Bor D, Stamatakis EA (2022) A synergistic core for human brain evolution and cognition. Nat Neurosci 25:771–782.

Marder E, Goaillard J-M (2006) Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci 7:563–574.

Markov NT et al. (2014) Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. J Comp Neurol 522:225–259.

Markov NT, Ercsey-Ravasz M, Essen DCV, Knoblauch K, Toroczkai Z, Kennedy H (2013) Cortical high-density counterstream architectures. Science 342:1238406.

Marr D, Poggio T (1976) From understanding computation to understanding neural circuitry. A.I. Memo 357. Cambridge, MA: Massachusetts Institute of Technology.

Maunsell JHR, Treue S (2006) Feature-based attention in visual cortex. Trends Neurosci 29:317–322.

McClelland JL, Lambón Ralph MA (2013) *Cognitive neurosciences: emergence of mind from brain*. London: Henry Stewart Talks.

McGill WJ (1954) Multivariate information transmission. Psychometrika 19: 97–116.

Mel BW, Fiser J (2000) Minimizing binding errors using learned conjunctive features. Neural Comput 12:731–762.

Miller KD (2016) Canonical computations of cerebral cortex. Curr Opin Neurobiol 37:75–84.

Muller L, Chavane F, Reynolds J, Sejnowski TJ (2018) Cortical travelling waves: mechanisms and computational principles. Nat Rev Neurosci 19: 255–268.

Nadim F, Brezina V, Destexhe A, Linster C (2008) State dependence of network output: modeling and experiments. J Neurosci 28:11806–11813.

Packard NH (1988) Adaptation toward the edge-of-chaos. In: *Dynamic patterns in complex systems* (Kelso JAS, Mandell AJ, Shlesinger MF, eds). Singapore: World Scientific.

Palmigiano A, Geisel T, Wolf F, Battaglia D (2017) Flexible information routing by transient synchrony. Nat Neurosci 20:1014–1022.

Paneri S, Gregoriou GG (2017) Top-down control of visual attention by the prefrontal cortex. Functional specialization and long-range interactions. Front Neurosci 11:545.

Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98:1064–1072.

Pedreschi N, Bernard C, Clawson W, Quilichini P, Barrat A, Battaglia D (2020) Dynamic core–periphery structure of information sharing networks in entorhinal cortex and hippocampus. Netw Neurosci 4:946–975.

Pica G, Soltanipour M, Panzeri S (2019) Using intersection information to map stimulus information transfer within neural networks. BioSystems 185:104028.

Pinzuti E, Wollstadt P, Gutknecht A, Tüscher O, Wibral M (2020) Measuring spectrally-resolved information transfer. PLoS Comput Biol 16:e1008526.

Pouille F, Scanziani M (2004) Routing of spike series by dynamic circuits in the hippocampus. Nature 429:717–723.

Price CJ (2018) The evolution of cognitive models: from neuropsychology to neuroimaging and back. Cortex 107:37–49.

Rosas FE, Mediano PA, Jensen HJ, Seth AK, Barrett AB, Carhart-Harris RL, Bor D (2020) Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. PLoS Comput Biol 16: e1008289.

Roxin A, Brunel N, Hansel D (2005) Role of delays in shaping spatiotemporal dynamics of neuronal activity in large networks. Phys Rev Lett 94:238103.

Roxin A, Brunel N, Hansel D (2006) Rate models with delays and the dynamics of large networks of spiking neurons. Prog Theor Phys Supp 161:68–85.

Rumelhart DE, McClelland JM (1986) *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Saban W, Raz G, Grabner RH, Gabay S, Kadosh RC (2021) Primitive visual channels have a causal role in cognitive transfer. Sci Rep 11:8759.

Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85:461–464.

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:623–656.

Shaw R (1984) *The dripping faucet as a model chaotic system*. Santa Cruz, CA: Aerial Press.

Shine JM, Bissett PG, Bell PT, Koyejo O, Balsters JH, Gorgolewski KJ, Moodie CA, Poldrack RA (2016) The dynamics of functional brain networks: integrated network states during cognitive task performance. Neuron 92:544–554.

Singley MK, Anderson JR (1989) *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Stetter M, Bartsch H, Obermayer K (2000) A mean-field model for orientation tuning, contrast saturation, and contextual effects in the primary visual cortex. Biol Cybernet 82:291–304.

Taatgen NA (2013) The nature and transfer of cognitive skills. Psychol Rev 120:439–471.

Treue S (2001) Neural correlates of attention in primate visual cortex. Trends Neurosci 24:295–300.

Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. Neural Comput 7:399–407.

Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy-a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci 30:45–67.

Voges N, Perrinet L (2012) Complex dynamics in recurrent cortical networks based on spatially realistic connectivities. Front Comput Neurosci 6:41.

Wibral M, Lizier JT, Vögler S, Priesemann V, Galuske R (2014) Local active information storage as a tool to understand distributed neural information processing. Front Neuroinform 8:1.

Wibral M, Priesemann V, Kay JW, Lizier JT, Phillips WA (2017) Partial information decomposition as a unified approach to the specification of neural goal functions. Brain Cogn 112:25–38.

Wickelgren WA (1999) Webs, cell assemblies, and chunking in neural nets: introduction. Can J Exp Psychol 53:118–131.

Wiener N (1956) The theory of prediction. In: *Modern mathematics for engineers* (Beckenbach E, ed), New York: McGraw-Hill.

Wilkes MV, Wheeler DJ, Gill SJ (1951) *The preparation of programs for an electronic digital computer*. Los Angeles, CA: Tomash Publishers.

Williams PL, Beer RD (2010) Nonnegative decomposition of multivariate information. arXiv:1004.2515 [cs.IT]. https://arxiv.org/abs/1004.2515.

Yger P, El Boustani S, Destexhe A, Frégnac Y (2011) Topologically invariant macroscopic statistics in balanced networks of conductance-based integrate-and-fire neurons. J Comput Neurosci 31:229–245.

Zylberberg A, Dehaene S, Roelfsema PR, Sigman M (2011) The human Turing machine: a neural framework for mental programs. Trends Cogn Sci 15:293–300.