Communication in large-scale brain networks: theory and application

by

Bratislav Mišić

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Psychology
University of Toronto

Copyright © 2014 by Bratislav Mišić
Abstract

Communication in large-scale brain networks: theory and application

Bratislav Mišić
Doctor of Philosophy
Graduate Department of Psychology
University of Toronto
2014

The complex connectivity of the cerebral cortex suggests that inter-regional communication is a primary function. The present body of work represents an effort to establish a conceptual and methodological framework in which to study integration and communication in brain networks.

Studies 1 and 2 describe a time series analysis method that utilizes time-frequency fluctuations of neural activity to infer how information is integrated by individual nodes of the network. Study 1 develops the method and demonstrates how it can be applied to resting-state EEG recordings. I show that the method is sensitive to changes in cognitive state and that it is capable of discriminating between eyes-open and eyes-closed recordings. In Study 2, the method is applied to a developmental MEG data set, in which participants performed a face processing task. Processing of upright faces was associated with a small number of sustained integrative episodes in right fusiform gyrus. When faces were inverted, face processing was fractured and associated with a larger number of shorter integrative stages.

Studies 3 and 4 describe a platform to simulate communication in large-scale brain networks. An anatomically realistic structural network was implemented as a dynamic system in which units of information flow along white-matter projections. Study 3 investigates how structural connectivity influences the efficacy of communication. I show that the topology of brain networks appears to be organized in a way that prioritizes speed of information transmission, rather than reliability and total throughput.
In addition, I demonstrate that the rich club of hub nodes is a critical topological feature that governs how signal traffic is distributed in the network. Study 4 explores the role of a particular node: subfield CA1 of the hippocampus. I show that, despite the fact that CA1 does not appear to be important from a topological perspective, in the context of networked communication it is one of the most important nodes in the brain and acts as a convergence zone for signal traffic.

Altogether, these data demonstrate the utility of models that embody the notion of communication and how such models offer novel and unique insight into brain structure and function.
Acknowledgements

From my first day, Randy has been a mentor and a friend. He has always treated me like a man and a colleague and given me the freedom to find my own path. I have learned many things from him, but perhaps the most important is that in research, you need to have a vision and the conviction to see it through. This thesis represents a highly unorthodox approach and no other advisor in the world would have trusted and encouraged me to pursue it. I hope that this work is one of many milestones for us.

Throughout my time in graduate school, I have had the opportunity to work with some of the best scientists in our field. I owe much to Olaf Sporns and Margot Taylor, both of whom I have worked with closely. I would also like to express my deepest gratitude to Bruce Schneider and Tomáš Paus, who have advised me and helped me at key times in my graduate career.

I have been fortunate to be part of the McLab at the same time as so many talented individuals. In graduate school, you learn from your labmates, you share aspirations and experiences, as well as frustrations (particularly about Reviewer no. 2). They teach you how to program, who to ask for help, when to push harder and when to let off. I extend my warmest thanks to all members of the lab, particularly Natasa, Andreea, Zee, Marc, Vasily, Michael, Kelly, Michele, Anjali, Erin, Sarah, Tanya, Maria, Grigori, Gleb, Hongye and Luiz. Getting up in the morning and going to the lab has never been a chore, but something that I look forward to because of the people there.

To pursue a graduate degree you need love and support at home. Hugs, kisses and chocolate are what makes life fun while you’re waiting for MATLAB to finish. Za Vecu, Lelu i Voju, volim vas najviše na svetu i ovo je, u najmanju ruku, isto toliko vaša zasluga koliko i moja. Mogu iskreno reći da sam našao zvanje u životu i da ste vi to omogućili. To Elena, I love you very much. This is for you.

The research conducted in this thesis was supported by a Canada Graduate Scholarship award from the Natural Sciences and Engineering Research Council of Canada, an Ontario Graduate Scholarship award from the Ontario Ministry of Training, Colleges and Universities, by the Wiseman Graduate Student Fellowship and Men’s Service Group Graduate Student Fellowship from Baycrest Hospital, by the Doctoral Completion Award from the Faculty of Arts and Science and by a grant from the JS McDonnell Foundation (JSMF22002082).
# Contents

1 General introduction ................................................. 1
   1.1 Motivation ..................................................... 1
   1.2 Network basis of brain function ............................... 3
   1.3 The graph model ................................................. 5
   1.4 Nonlinear dynamic systems models .............................. 7
   1.5 Queueing systems ............................................... 9
   1.6 Linking queueing systems and brain networks ................. 11
   1.7 Overview of studies .............................................. 12

2 Inter-event times in human EEG ................................. 15
   2.1 Introduction ..................................................... 15
   2.2 Materials and methods .......................................... 17
      2.2.1 Electroencephalogram (EEG) acquisition .................. 17
      2.2.2 Wavelet transform .......................................... 18
      2.2.3 Inter-event time distributions .............................. 19
      2.2.4 Gamma distribution ......................................... 23
      2.2.5 Partial least-squares (PLS) analysis ....................... 26
   2.3 Results .......................................................... 29
      2.3.1 Inter-event time distributions .............................. 29
      2.3.2 Shape parameter $k$ ......................................... 30
      2.3.3 Scale parameter $\theta$ ...................................... 31
   2.4 Discussion ....................................................... 32
      2.4.1 Importance of the gamma distribution .................... 33
      2.4.2 Methodological considerations ............................. 35
      2.4.3 Conclusion .................................................. 36
# 3 Inter-event times and face processing

## 3.1 Introduction

## 3.2 Materials and methods

### 3.2.1 Participants

### 3.2.2 Stimuli and task

### 3.2.3 Magnetoencephalogram (MEG) acquisition

### 3.2.4 Event-related beamforming (ERB)

### 3.2.5 Wavelet transform

### 3.2.6 Inter-event time distributions

### 3.2.7 Partial least-squares (PLS) analysis

## 3.3 Results

### 3.3.1 Inter-event time distributions

### 3.3.2 Development

### 3.3.3 Development and face inversion

## 3.4 Discussion

### 3.4.1 Inter-event time distributions

### 3.4.2 Development

### 3.4.3 Development and face inversion

### 3.4.4 Methodological considerations

### 3.4.5 Conclusion

# 4 Communication efficiency in brain networks

## 4.1 Introduction

## 4.2 Materials and methods

### 4.2.1 Anatomical and reference networks

### 4.2.2 Rich club detection

### 4.2.3 Discrete-event simulation

## 4.3 Results

### 4.3.1 Network statistics

### 4.3.2 Node statistics

### 4.3.3 Edge statistics

### 4.3.4 Path statistics

## 4.4 Discussion
List of Tables

C.1 Comparing scenarios .................................................. 114
F.1 Network comparisons for the Transit Time statistic .................. 123
F.2 Network comparisons for the Throughput statistic .................... 123
F.3 Comparisons for the node contents statistic ........................... 124
H.1 Rich club area names ................................................ 129
# List of Figures

2.1 Wavelet transform ................................................. 20  
2.2 Peak picking ......................................................... 22  
2.3 Overview of processing pipeline ................................. 23  
2.4 Gamma distribution .................................................. 24  
2.5 Empirical probability distributions .............................. 25  
2.6 Gamma as a sum of exponentials ................................. 27  
2.7 Group means for the shape parameter ........................... 29  
2.8 Group means for the scale parameter ........................... 30  
2.9 Bootstrap ratio maps ............................................... 31  
3.1 Source locations ..................................................... 42  
3.2 Processing pipeline ................................................ 44  
3.3 Individual inter-event time distributions ........................ 46  
3.4 Developmental effect captured by the scale parameter ....... 48  
3.5 Developmental effect captured by the shape parameter ....... 49  
3.6 Face inversion effect captured by the scale parameter ...... 51  
3.7 Face inversion effect captured by the shape parameter ...... 52  
4.1 Discrete-event simulation .......................................... 60  
4.2 Small-world network ............................................... 62  
4.3 System-level statistics .............................................. 68  
4.4 Node-level statistics ................................................ 70  
4.5 Node contents for the CoCoMac and randomized control networks .... 71  
4.6 Edge throughput ..................................................... 72  
4.7 Delays and deliveries ............................................... 73  
5.1 Node metrics ......................................................... 82
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Role of network topology</td>
<td>83</td>
</tr>
<tr>
<td>5.3 CA1 as a communication outlier</td>
<td>85</td>
</tr>
<tr>
<td>5.4 Assortativity</td>
<td>86</td>
</tr>
<tr>
<td>5.5 Degree imbalances</td>
<td>87</td>
</tr>
<tr>
<td>5.6 Neighbourhood of CA1</td>
<td>88</td>
</tr>
<tr>
<td>5.7 Role of directionality</td>
<td>89</td>
</tr>
<tr>
<td>A.1 System statistics: analytical and numerical results</td>
<td>107</td>
</tr>
<tr>
<td>A.2 Node statistics: analytical and numerical results</td>
<td>108</td>
</tr>
<tr>
<td>B.1 Effect of buffer size</td>
<td>111</td>
</tr>
<tr>
<td>D.1 Effect of interpolation</td>
<td>117</td>
</tr>
<tr>
<td>E.1 Effect of rich club size</td>
<td>120</td>
</tr>
<tr>
<td>G.1 Effect of increasing simulation intensity</td>
<td>126</td>
</tr>
<tr>
<td>H.1 Rich club of the macaque network</td>
<td>128</td>
</tr>
</tbody>
</table>
This here game is more than the rep you carry, the corner you hold. You gotta be fierce, I know that, but more than that, you gotta show some flex, give and take on both sides.

“Stringer” Bell
Chapter 1

General introduction

1.1 Motivation

The axiomatic view of human psychology is that cognition is limited and this is reflected by a finite capacity for various faculties, such as attention and working memory. For instance, iconic memory is fleeting, barely lasting more than a few moments (Sperling 1960). A typical high-functioning human being is only able to maintain approximately seven items in working memory (Miller 1956). Even the ability to simply register two successive events presents considerable difficulties and eventually becomes impossible if they occur within a short span of each other (Raymond et al. 1992). However, the reasons for such limits are unknown and typically attributed to scarce “cognitive resources”.

The topology of large-scale structural and functional human brain networks confers a capacity for information processing far greater than comparable networks with similar size and connection density. During brain development, extensive myelination and synaptic pruning serve to establish an optimal balance between local segregation and global integration, leading to a system with high complexity (Tononi et al. 1994). This balance is mediated by the emergence of highly connected regions that integrate disparate modules (Sporns et al. 2007; Zamora-López et al. 2010; Fair et al. 2009). Neural infrastructure obeys several generic organizational principles ubiquitous in biological systems (Bullmore et al. 2009), including small-world connectivity (Sporns and Zwi 2004; Bassett and Bullmore 2006).

By the same token, network topology could also constrain cognitive capacity. Just as some brain areas bridge distant communities and function as hubs by virtue of their
connectivity, other areas may be particularly ill-suited as conduits for information transfer. Under conditions of elevated network traffic, how is signal traffic distributed on brain networks? How does the connectivity of the brain networks influence the time it takes for information to travel? Are there areas that are particularly pivotal for the exchange of information? How is information processing reflected by neural activity and how does information processing change in response to changes in external demand?

Such questions are of much interest in various types of communication systems – from telephony to the Internet – and are the subject of a branch of probability theory known as queueing analysis (Kleinrock 1976a,b; Takagi 1991b; Nelson 1995; Bolch et al. 2001). Units of information (sometimes referred to as “packets” or “signal units”) are routed through the network and processed at each node they traverse for some stochastic duration. If a node is busy when a signal unit arrives to it, a queue is formed. Therefore, queueing networks are used for modeling and performance evaluation. They provide a framework to generate numerous system statistics, such as utilization of nodes and transit times for signal units.

In the following thesis I explore how network connectivity supports information processing and communication. I argue that the configuration of large-scale brain networks need not be seen as an abashed advantage, but rather as a compromise among many different constraints that limit all aspects of neural function, including perception, cognition and action. In particular, I highlight two areas of interest: (1) how information is assimilated by individual nodes and (2) how information is communicated between nodes in the network. I hypothesize that network and time series analysis from the perspective of queueing theory may be an ideal way to quantify information flow in the brain and to extract useful performance statistics. In the following section, I discuss how critical features of brain network organization – suggested by recent studies using graph theoretical analysis and nonlinear dynamic systems models – may be implemented in a physiologically realistic queueing network model of the brain, including realistic connectivity, noise and time-delayed transmission.
1.2 Network basis of brain function

A fundamental challenge in cognitive neuroscience is to understand how neural operations engender mental phenomena, from sensation to perception to higher cognition. Early studies, looking at cognitive deficits following lesions, often attributed discrete mental operations to specific brain regions. Modern neuroimaging techniques, which can simultaneously measure activity in the whole brain during some mental function, have given rise to neurobiological theories that emphasize the distributed and integrative nature of neural operations. Brain regions do not operate in isolation, but as part of an interacting network (McIntosh et al. 1999; Stephan et al. 2003), and the contribution of a particular region must depend on the status of the other elements of the network, a principle known as “neural context” (McIntosh 2000; McIntosh 1999; Bressler and McIntosh 2007).

Neural interactions can be thought of in terms of structure-function relationships, whereby the anatomical connectivity between neural elements provides the platform for the exchange of information. In neuroimaging, neural interactions are conceptualized and estimated in terms of two distinct but related notions of connectivity. Functional connectivity refers to temporal associations between neural elements and may be estimated using correlation or covariance, as well as coherence and mutual information. Thus, two elements with statistical interdependency are said to be functionally connected. Yet a functional connection between two neural elements does not imply that they are communicating directly, as their covariation may be due to common inputs from another source. Effective connectivity is a logical progression from functional connectivity, and refers to the direct influence of one neural element on another (Aertsen et al. 1989; Friston 1994).

A unifying framework in brain network analysis is the notion of simultaneous segregation and integration (Tononi et al. 1994; Sporns et al. 2000). Segregation refers to the localization of function: the fact that neurons tend to be organized into groups, characterized by common inputs, outputs and response properties. As a result, proximal units tend to specialize for similar functions at multiple spatial scales, from minicolumns to large-scale brain “regions”. This notion is further supported by lesion studies, which show that focal lesions tend to produce specific functional deficits. Likewise, areas tend to preferentially respond to certain types of stimuli or external demands. This suggests that there exist specialized sub-systems that perform spe-
specific functions with relatively little overlap (e.g. a specialized visual system, auditory system, etc.).

Integration refers to the fact that communication between specialized neural elements is necessary to produce a coherent, unified experience. The architecture of the brain ensures that any two neurons are separated by only a small number of synapses, presumably to promote the exchange of information. Likewise, the ubiquity of reciprocal connections facilitates massively parallel processing and is reflected by widespread temporal correlations between neural elements (i.e. functional connectivity). The existence of polysensory association areas, mainly in neocortex, that respond to complex, multimodal stimuli, further suggests that a basic function of brain networks is to assimilate information from distributed sources.

In neuroimaging, segregation and integration have traditionally been operationalized in terms of activation and functional or effective connectivity. In activation studies, changes in neural activity due to some experimental manipulation are used to infer which parts of the brain support a particular function. In functional and effective connectivity studies, changes in temporal associations between neural elements are used to infer whether the exchange of information between those elements supports a particular function. In addition, effective connectivity models permit inference on the directionality of causal influence.

However, a primary function of brain networks is to assimilate and communicate information. Studies typically focus on increases and decreases in neural activity and connectivity, rather than information flow and information processing per se. Neuroimaging techniques, including functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG) and positron emission tomography (PET) all measure biophysical changes in the brain rather than information processing per se. Therefore, constructs such as communication necessitate some form of model and information processing must be inferred rather than directly measured from the data. The goal of this thesis is to operationalize information processing in brain networks and to create models that can be used to estimate various aspects of communication.

In the following two sections I give a brief overview of two classes of paradigms that have emerged in the literature to study the functional capacity of large-scale brain networks. The graph model allows us to articulate and formally study the network organization of the brain. Dynamic systems models allow us to simulate neural
activity on a realistic structural substrate. To foreshadow, although both paradigms have proven remarkably useful, they involve models that do not embody the notion of communication - how information is exchanged between distributed elements.

1.3 The graph model

The graph model has become widely used to describe and quantify the topology of cortical networks (Bullmore and Sporns 2009, 2011; Rubinov and Sporns 2010). Under this framework, the brain is spatially discretized into a set of regional nodes that are interconnected by a set of edges. In structural brain networks the edges correspond to white matter projections, whereas in functional networks they represent some measure of association, such as a Pearson correlation coefficient, mutual information or transfer entropy. The former can be delineated using chemical tracers (Stephan et al. 2001; Kötter 2004) or diffusion-weighted magnetic resonance imaging (MRI) (Iturria-Medina et al. 2007, 2008; Gong et al. 2009; Hagmann et al. 2007, 2008), while the latter can be derived using virtually any measure of neural activity, such as functional MRI (fMRI) (Achard et al. 2006; Bassett et al. 2008; Buckner et al. 2009), EEG (Fallani et al. 2007a,b; Mišić et al. 2011b) or MEG (Bassett et al. 2006, 2009).

Relative to reference networks with comparable connection density, such as a regular lattice or a random network, networks extracted from neuroimaging studies possess a whole host of properties that make them ideally suited for information processing. Perhaps the most important is the ability to reconcile specialized processing in local neighborhoods (segregation) with distributed processing over multiple, otherwise independent domains (integration), which reflects the functional complexity of the system (Tononi et al. 1994). As previously mentioned, in the human brain the coexistence of functional segregation and integration is made possible by the underlying anatomy, which is characterized by dense local connectivity and sparse distal connectivity (Tononi et al. 1994; Sporns et al. 2000).

Small-world networks are a salient and well-studied example of such organization (Watts and Strogatz 1998). In this class of networks, high clustering or “cliquishness” is offset by a small number of long-range connections. The end result is a global topology in which the expected minimum path between any two points is considerably shorter than in a regular network with the same number of nodes and edges. This type of
organization is well-documented in structural brain networks (Bassett and Bullmore 2006; Hilgetag et al. 2000a; Sporns and Zwi 2004). The ease with which small-world networks can be traversed potentially explains their ubiquity in nature, from protein-protein interactions to nervous systems to social networks (Watts and Strogatz 1998).

Large-scale brain networks are inherently segregated into communities, clusters or modules. Nodes within a module tend to be densely interconnected, sharing common inputs and outputs. As a result, members of a particular module will often have similar response properties and common function. The modular organization of brain networks spans multiple spatial scales, resulting in a hierarchy of communities that correspond to distinct functional sub-divisions (Bassett et al. 2008). This facilitates specialized processing across a variety of systems and sub-systems. For example, the primate brain can be clustered into visual, auditory, somatosensory-motor and frontolimbic areas, while the visual cluster can be further divided into smaller clusters that correspond to the dorsal and ventral visual streams (Hilgetag et al. 2000b).

The integration of the brain during development is facilitated by the emergence of a set of regions with disproportionately dense connectivity profiles (Fair et al. 2009; Hagmann et al. 2010). These regions act as “hubs” in the network, occupying strategic positions along the shortest paths between other pairs of nodes. Situated along the midline in frontal and parietal cortices, hubs form the structural “core” of the brain network and in a sense they are the fulcrum of the entire system (Hagmann et al. 2008). The presence of hubs ensures that the effective communication distance between nodes is relatively short. Their integrative influence is exerted both locally to define modules and bind them together (“provincial” hubs), as well as globally, to bridge distant communities (“connector” hubs) (Sporns et al. 2007). Therefore, hubs are hypothesized to streamline parallel processing and increase efficiency relative to metabolic and wiring cost (Achard and Bullmore 2007; Bassett et al. 2009). Recent studies have also uncovered a “rich club” of hub nodes that are more densely interconnected with each other than predicted by chance (Zamora-López et al. 2010; van den Heuvel and Sporns 2011) and that participate in a disproportionately high number of shortest paths in the network (van den Heuvel et al. 2012; Harriger et al. 2012). The rich club is hypothesized to act as a central backbone for signal traffic, allowing for rapid integration and dissemination of signal traffic (van den Heuvel et al. 2012).

The infrastructure of the cerebral cortex shapes and ultimately constrains the resultant expression of functional networks (Honey et al. 2009, 2010). The topological
attributes of functional networks largely mirror those of the structural networks from which they arise. In terms of organization, functional networks also display small-world connectivity (Achard et al. 2006; Bassett et al. 2006; Salvador et al. 2005; Stam 2004). Regions that are anatomical hubs also tend to be functional hubs (Hagmann et al. 2008; Honey et al. 2009; Hagmann et al. 2010; Tomasi and Volkow 2010). The link between structural and functional connectivity is an important one: functional associations engendered by structural connections are thought to reflect the prevailing flow of information.

Considerable effort has been directed toward understanding how functional capabilities change if the underlying anatomical network is significantly altered in some fashion, such as by the removal of nodes and edges (Achard et al. 2006; Honey and Sporns 2008; Alstott et al. 2009; Rubinov et al. 2009b). These studies mimic the effect of structural insult, such as lesions associated with stroke or dysconnectivity associated with schizophrenia. A consistent finding has been that networks with brain-like connectivity are more resilient to deletions than reference networks. All factors being equal, core characteristics of brain networks persevere even under heavy losses. The notion that some nodes are more important than others is reinforced by these types of analyses. Systematic or “targeted” deletions of central nodes are associated with far more deleterious consequences than deletions of random nodes.

1.4 Nonlinear dynamic systems models

The structural connectivity of the brain provides the platform for emergent psychological phenomena. In this sense, the graph model is a natural framework in which to study the functional interdependencies fostered by physiological connectivity. An emerging approach has been to superimpose a nonlinear dynamical system over a realistic brain-like anatomical foundation (Sporns et al. 2000). The holistic expression of the simulated brain is thus generated by synergetic action among the connected neuronal ensembles (Stam et al. 2003). Typically, the dynamics of each node represent the mean electrical potential of a small population of neurons, termed “neural masses”. Their temporal evolution is described by a system of stochastic nonlinear differential equations, while interdependencies between the nodes are instilled by allowing connected nodes to exert some influence on each other (“coupling”). For reviews on this
topic I refer the reader to (Deco et al. 2008) and (Breakspear et al. 2010). The objective here is to describe the features of such large-scale models that have been deemed essential for producing realistic dynamics.

A key observation in empirical studies is that functional connectivity patterns recovered from neural activity tend to recapitulate the underlying structural connectivity, but only to a limited extent (Honey et al. 2009). Although structural connectivity largely constrains functional connectivity and represents the dominant configuration, the spatiotemporal dynamics of the system are prone to occasional, entirely spontaneous forays into other configurations. It is hypothesized that this propensity to explore alternate states confers a high degree of flexibility and expands the dynamic répertoire of the brain (Ghosh et al. 2008b,a; Deco et al. 2009; Jirsa et al. 2010). Nonlinear dynamical models are able to reproduce this behaviour, with functional patterns largely conforming to the structural foundation over long time scales but temporarily straying at short time scales (Honey et al. 2007). Moreover, ongoing oscillations in such models tend to detach into distinct and anticorrelated functional clusters (Honey et al. 2007; Ghosh et al. 2008b; Deco et al. 2009), closely mirroring the mutual antagonism observed in neural activity recorded from the resting brain (Fox et al. 2005).

This self-organizing behavior arises from the interplay between the local and global dynamics and is made possible by the space-time structure embodied by interregional couplings of the network (Ghosh et al. 2008b; Jirsa et al. 2010). First and foremost, the characteristic topology of brain networks is critical to these emergent properties (Honey et al. 2007; Ghosh et al. 2008b; Honey et al. 2009). The varying coupling strengths and transmission delays between regions are the defining attributes of the network and enhance the state space of the brain’s dynamics (Honey et al. 2007; Ghosh et al. 2008a,b; Deco et al. 2009). The tendency to exercise alternative options and reconfigure spontaneously is enabled by the presence of a moderate level of local noise (Ghosh et al. 2008b; Deco et al. 2009). As the brain typically occupies a high-energy state – at the “edge of instability” – noisy fluctuations serve to induce transitions and allow the system to traverse the state space.
1.5 Queueing systems

The graph model has demonstrated that the infrastructure of cortical networks confers an adaptive advantage. Namely, the distributed organization of the cerebral cortex is optimized for information processing and presumably makes neural function more efficient. Empirical studies and dynamic systems models have identified system parameters that are critical for neural function, including topology, time delays and noise.

The graph-theoretic approach emphasizes the static connectivity between neural elements, while the nonlinear dynamics approach models how this connectivity interacts with the biophysical properties of neural elements to produce the global patterns of neural activity observed in functional neuroimaging. However, neither the graph theoretic metrics applied to static structural and functional networks nor dynamic systems models superimposed on such networks have been able to characterize the relationship between cortical connectivity and information processing. How is information communicated on the network and integrated by individual nodes? Which regions are bottlenecks for information processing? Which subnetworks are vulnerable to congestion? How does the flux of information through individual nodes relate to network conditions? There remains the need for a theoretical framework to conceptualize and analyze network traffic at the large spatial scale.

Questions of this type are the domain of queueing analysis, a branch of probability theory that seeks to mathematically capture both the transient and long-term behavior of a system with waiting lines, or queues. The basic architectural unit of a queueing network is the service station or server, which is located at each node or brain region. Units of information (also known as packets, signals, messages or customers, depending on the application) arrive at each node and are processed immediately for some period of time if the node is free. If the signal unit arrives to a busy node, a queue is formed. Upon completion of service, the signal unit is transmitted to another node in the network. Possible paths for each signal unit are defined by the graph itself.

The key facet of the system is that both the sequence of arrivals to and service times for each node may be stochastic processes. Arrivals may originate from inside or outside the network and the arrival process is specified by a probability distribution of inter-arrival times. Likewise, service times are also described by some probability distribution. The manner in which signal units get routed through the network is
governed by a switching process.

What do signal units and servers represent in the brain and how could such a seemingly abstract model be physiologically realistic? Here I make two important assumptions. The first is that neural elements communicate and send information to each other. Communication may take place via action potential spikes, spike trains or volleys of spike trains. A signal unit therefore does not represent any specific event, but rather, the ability of one neural element to influence another. The second assumption is that neural elements process or transform signal units in some manner. Thus, signal units spend some finite period of time at each node they visit, rather than passing through instantaneously.

A queueing framework can capture several key aspects of neural biology. Realistic structural connectivity from tract tracing or diffusion-weighted imaging can easily be incorporated by placing servers at each node in the network and restricting the travel of signal units to existing edges. The entire space-time structure of the cortical network can be brought to bear on the flow of information by additionally incorporating time delays. Signal units that have completed service are not instantaneously forwarded to the next node, but take time to traverse the intervening edge. Delay time is determined according to the length of the pathway (which can be approximated by the Euclidean distance between the nodes if their position in three-dimensional space is known) and the conduction velocity of axons in the pathway. Thus, local dynamics are embedded in the global brain network and their mode of operation is inseparable from the greater system-level context.

The potency of action potentials arriving at a synapse is not maintained indefinitely. Unless a post-synaptic potential is triggered immediately, the cell membrane gradually reverts back to the resting potential and the effect of the original action potential is swiftly diminished. This suggests that the efficacy of information transfer in the brain is not uniform across time, but is biased in favor of the most recently-arriving information. This property can be simulated by setting the queueing discipline to last-come-first-served (LCFS), such that signal units arriving to a busy server immediately move to the front of the queue. This solution is attractive for several reasons. First, like the more common first-come-first-served (FCFS) queueing discipline, it is also impartial and avoids problematic notions like priorities and multiple classes of signal units. Second, most of the important system statistics like utilization and probability of blocking will be identical to a network with FCFS discipline and only the mean
waiting times will be substantially different under the two schemes (Kleinrock 1976a).

The stochastic nature of queueing networks is ideally suited to mimic real-life variability and noise in biological systems such as the brain. For example, neural transmission is marked by low fidelity and is subject both to stochastic fluctuations due to imprecise timing of cellular processes as well as loss of information (Faisal et al. 2008). In a queueing network, this imprecision is reflected in the fact that processing times at each brain region are not deterministic and instead are drawn from a probability distribution. Moreover, this distribution need not be assumed but can be estimated from real neurophysiological data. Loss of information is made possible by imposing limits on buffer capacity. If buffer capacity is exceeded either the earliest or latest-arriving packet is lost, depending on the queueing discipline.

1.6 Linking queueing systems and brain networks

The extension of the queueing framework to brain networks would permit the derivation of a whole host of novel measures to characterize the integration and communication in the brain. In particular, duration metrics are an important aspect of queueing systems. At the level of individual nodes, the queueing time and service time are informative. If the arrival process and the output process of a node is known, the manner in which signal units are serviced can be inferred. If a signal unit has a specific source and destination, then the total transit time is of interest. If buffer capacity is finite, then “blocking” occurs (by analogy with blocked calls in telephony), where a signal unit is either dropped from a queue or displaced by another signal unit, depending on the queue discipline. The probability of blocking is an informative performance index for a system with finite capacity. Perhaps the most important of these statistics is the proportion of time the node is busy (utilization). In addition, the average lengths of busy and idle periods and the manner in which they alternate all reflect the overall dynamics and efficiency of the network.

What types of questions and inquiries would this approach facilitate? For individual neural elements, it may be possible to infer how information is assimilated and how integrative dynamics change due to experimental manipulations. From the perspective of network analysis, system statistics generated by a queueing model may be used to identify brain regions, pathways and subnetworks that are likely to expe-
Chapter 1. General introduction

Experience high throughput by virtue of their embedding in the structural network of the brain. The goal is to generate meaningful indices for the flow of information through individual nodes and the entire network.

Queueing models have already received some attention in cognitive psychology (for an excellent review, see (Liu 1996)). However, these models were formulated to reproduce specific phenomena such as reaction times (Wu and Liu 2008) and latencies of neurophysiological responses (e.g. the P300 evoked potential) (Wu et al. 2008). The approach explored in this thesis is substantially different in scope and is geared towards generating descriptive, dependent measures about communication. In addition, these prior models abstract away the structural connectivity of the brain in favor of a smaller number of structural compartments, such as distinct “perceptual”, “cognitive” and “motor” modules. The connectivity in such models is not physiologically realistic in the sense that it is not directly measured using structural imaging and key features such as sparseness, small-worldness and degree sequences are not preserved. Similarly, realistic time delays in signal unit travel are usually not incorporated.

For mathematical convenience, prior studies have often assumed that service and inter-arrival times are exponentially distributed. As I show in Chapters 2 and 3, these assumptions may be too strong and possibly invalid. In addition, previous queueing models of the brain assumed infinite buffer capacity and a FCFS queueing discipline, eschewing the possibility of short memory and imperfect fidelity. As we argued above, these assumptions are not realistic and must be changed in order to accurately estimate blocking probabilities and transit times. Moreover, most such queueing models implement deterministic routing, which cannot capture the fluid reconfiguration of functional networks and is unlikely, even from a theoretical point of view (Graham and Rockmore 2010).

1.7 Overview of studies

The present body of work represents an effort to establish a conceptual and methodological framework in which to study integration and communication in brain networks. I use a queueing theoretic approach as a logical extension of the dynamic systems and graph theoretic paradigms. The overall purpose of the two methods presented in this thesis is to use queueing network models to generate descriptive
statistics about information flow in large-scale brain networks.

The first two studies (Chapters 2 and 3) focus on an empirical method that utilizes time-frequency fluctuations of neural activity to infer how information is integrated by individual neural elements. I show that a queueing-theoretic generative model provides a good fit to the empirical data and allows novel insight into how information is assimilated by neural elements. Study 1 (Chapter 2) develops the method and – as proof of principle – demonstrates how it can be applied to resting-state EEG recordings. Importantly, I show that the method is sensitive to changes in cognitive state and is capable of discriminating between eyes-open and eyes-closed resting state recordings. Study 2 (Chapter 3) demonstrates that the method is sensitive to development and cognitive engagement, and can be used to gain novel insight into cognitive development. I apply the method to a developmental MEG data set in which participants from 6 to 41 years of age performed a face processing task. In addition, I adapt the method to a more traditional cognitive experiment, with multiple shorter epochs.

The next two studies (Chapters 4 and 5) focus on a large-scale simulation of communication in brain networks. A physiologically realistic anatomical network was implemented as a dynamic system in which units of information flow along white-matter projections. Study 3 (Chapter 4) investigates how structural connectivity influences the efficacy of communication. I show that the topology of brain networks appears to be organized in a way that prioritizes speed of information transmission, rather than reliability and total throughput. In addition, I demonstrate that the rich club of hub nodes is a critical topological feature for network-wide communication. Study 4 (Chapter 5) explores the role of a particular node: CA1 of the hippocampus. I show that, despite the fact that CA1 does not appear to be important from a topological perspective, in the context of networked communication it is one of the most important nodes in the brain. By analyzing the communication patterns around CA1, I show that this is in large part due to a severe degree imbalance at various parahippocampal structures neighbouring CA1, which are known to have more afferent than efferent projections. The appendices contain further clarification with respect to the methods employed in Chapters 4 and 5.
The studies presented in this thesis were adapted from articles published, submitted or in preparation for submission. In order of the chapters presented, these articles are:


Chapter 5  Mišić B, Goni J, Sporns O and McIntosh AR A network convergence zone in the hippocampus. in preparation.
Chapter 2

Inter-event times in human EEG

adapted from Mišić et al. (2011) PLoS Comput Biol;
McIntosh & Mišić (2012) Annu Rev Psychol

2.1 Introduction

Recent years have witnessed a remarkable drive to characterize the large-scale structural topology of the brain. The graph model of cortical connectivity – whereby space is discretized and the brain is delineated as a set of regional nodes interconnected by white matter edges – has enabled the application of a whole host of network metrics (Bullmore and Sporns 2009; Rubinov and Sporns 2010). The cerebral connectome (Sporns et al. 2005) has been found to possess highly nontrivial properties that do not appear in random networks with comparable connection density and could potentially endow it with a greater capacity to process information. These include small-worldness (Hilgetag et al. 2000a; Sporns and Zwi 2004; Bassett and Bullmore 2006) and the presence of hubs (Sporns et al. 2007; Hagmann et al. 2008).

However, the functional consequences of this structural foundation are less clear and in general the translation from structure to function has been more difficult to understand. The emergent functional connectome has hitherto been studied by applying similar network analytic measures to graphs extracted from functional data. One approach has been to use these indices as a basis of comparison between networks defined by structural and functional connections. For example, physical links between nodes certainly beget sustained functional interactions and as a result functional brain networks map onto the underlying structural architecture to a great extent (Honey
et al. 2009; Hagmann et al. 2008; Honey et al. 2010). Another approach has been to study functional networks exclusively and without explicit reference to the underlying structural networks (Achard et al. 2006; Bassett et al. 2006).

An important aspect of brain network organization that remains to be investigated is the throughput of information at individual nodes. How does the flux of information vary across regions and under changing external and internal conditions? Do all nodes receive, process and relay messages at the same rate? Questions of this type often arise in relation to many classes of distributed communication networks (Kleinrock 1976a; Takagi 1991b; Bolch et al. 2001). Indeed, the brain must engage in networked computation (Bressler 1995; McIntosh 1999; McIntosh 2000), a challenge common to multiple types of telecommunication systems (Kurose and Ross 2008). Therefore, it may be possible to learn more about the functional architecture and organizational principles of the brain by treating it as a network of regions that emit units of information.

The present study represents the first step in adapting tools from telecommunications research to the problems in neuroscience. Namely, I show how electrophysiological recordings can be plausibly translated into a trace of departing units of information or output “events” and analyzed from the perspective of a telecommunication system. By casting the problem in this light, it may be possible to find new ways to describe, quantify and model the flow of information along the distributed brain network. One of the fundamental system statistics for modeling and performance evaluation of communication networks is the distribution of time between successive output events at each node (Kleinrock 1976a,b; Takagi 1991b; Nelson 1995; Bolch et al. 2001). The inter-event time depends on how messages get processed as well as the nature of their aggregated arrivals to a node and as such it reflects the flux of information through the network. In the present study I devised a method to delineate units of information in gross neurophysiological recordings and to fully characterize the distribution of their inter-event times.

I first describe an intuitive signal processing approach that can be used to extract such events from the electroencephalogram (EEG). Participants were at rest, with both eyes-open and eyes-closed conditions. The data were resolved in the time-frequency domain using a wavelet transform. Output events were defined as the local maxima in the EEG scalogram, a definition based on the direct physiological interpretation of the EEG. Peaks and bursts in EEG signal power represent the synchronous firing of post-synaptic potentials from a population of neurons. If we take the neuron soma to
be grey matter nodes in the network (as the graph model does), then the propagation of post-synaptic potentials to the axon hillock and along the axon may be thought of as the departure of a message. Thus, each peak in signal power is an event that marks the point in time at which a unit of information departs from that population of neurons. I show that the distribution of time between successive events (the inter-event time) is well described by the family of two-parameter gamma distributions. These distributions were fitted at each electrode and the two estimated parameters were then treated as dependent variables of neural activity.

If such events do indeed capture some aspect of information flow in brain networks, then several testable predictions can be made. First, the actual paths and sequences of “hops” between nodes will be largely determined by their structural connectivity, so inter-event time statistics should be region specific and their spatial distribution should be heterogeneous. Second, as external demands change, so too should the manner in which units of information are emitted across the network and the distribution of inter-event times at individual nodes should also be task-dependent. In particular, we expected the greatest change to be observed at or near occipital channels, given that the biggest difference between the eyes-closed and eyes-open states is the presence of visual input.

2.2 Materials and methods

2.2.1 Electroencephalogram (EEG) acquisition

The experimental protocol was approved by the Research Ethics Board of the Montreal Neurological Institute and Hospital. Fifty-six (29 male) healthy children 10 years old (mean 10.0, standard deviation 0.393 years) participated in the study (see (Poulsen et al. 2009) for details). The participants were asked to keep their eyes open or closed in 8 alternating 30 s epochs (4 each). The electroencephalogram (EEG) was continuously recorded from 128 scalp locations using a HydroCel geodesic sensor net (Electrical Geodesics, Inc., Eugene, OR) referenced to the vertex (Cz). The signal was digitized at a rate of 500 Hz. Impedances did not exceed 60 kΩ. All offline signal processing and artifact correction was performed using the EEGLAB toolbox (Delorme and Makeig 2004) for MATLAB (Mathworks, Inc.). Data were then average-referenced, digitally filtered [band-pass: 0.5-55 Hz; notch: 60 Hz] and epoched into 30 s segments. Only
the middle 20 s of each epoch (5-25 s) were used in the analysis to avoid excessive contamination associated with opening and closing of the eyes. In the absence of a true baseline, the temporal mean was subtracted from each epoch. Ocular (blinks and lateral eye movements) and muscle artifacts were identified and subtracted on a subject-by-subject basis using the Infomax independent components analysis (ICA) algorithm (Bell and Sejnowski 1995) implemented in EEGLAB.

2.2.2 Wavelet transform

Dynamic spectral changes were estimated using a wavelet transform (Torrence and Compo 1998), implemented in the Wavelet Toolbox for MATLAB (Mathworks, Inc.). Trial epochs were convolved with a complex Morlet wavelet in a sliding window and signal power was estimated as the modulus squared of the real-valued wavelet coefficients (Fig. 2.3B). The Morlet wavelet is a Gaussian-modulated complex sinusoid, so it is considered biologically plausible because it is more sensitive to transients in time series (more so than the windowed Fourier transform) and is widely used as an alternative way to model signals such as the EEG (Roach and Mathalon 2008). The mother wavelet had center frequency \( F_c \) equal to 1 Hz and envelope bandwidth \( F_b \), corresponding to the standard deviation) equal to 2 s. The wavelet can be described by the following expression:

\[
\Psi(t) = \sqrt{\frac{1}{\pi F_b}} \exp(2\pi i F_c t) \exp\left(-\frac{t^2}{F_b}\right),
\]

where \( i \) is an imaginary number. The continuous wavelet transform \( \omega(t) \) is obtained by temporal convolution of the wavelet \( \Psi(t) \) with a section of the signal \( m(t) \)

\[
\omega(t) = m(t) \otimes \Psi(t) = \int_{-\infty}^{\infty} m(\tau) \frac{1}{\sqrt{a}} \Psi^*(t-\tau) \frac{1}{a} \mathrm{d}\tau.
\]

where \( \otimes \) denotes the convolution operator, * denotes the complex conjugate, \( \tau \in \mathbb{R} \) denotes a translational value and \( a > 0 \) denotes the scale factor.

Fig. 2.1A illustrates how the wavelet transform is computed. A wavelet is initially convolved with a section at the start of the signal (step 1, Eq. 2.2). If the wavelet is complex, the convolution produces complex-valued wavelet coefficients, which are analogous to correlation coefficients in the sense that they reflect the degree of similarity between the wavelet and that section of the signal. The wavelet coefficients
consist of both a real and imaginary component, which reflect the amplitude of the oscillation. The angle that this vector forms with the real axis reflects the phase of the oscillation. The wavelet is then shifted along the signal (by $\tau$) and the wavelet coefficients are re-computed (step 2). Once the whole signal has been covered, the wavelet is scaled (compressed or dillated, by $a$) and the same procedure is performed (step 3).

The shifting of the wavelet by the translational value $\tau$ simply delays or advances its onset. The scaling of the wavelet by the factor $a$ affects the frequency to which the wavelet is sensitive and this is illustrated in Fig. 2.1B. Here, to isolate a time series of wavelet coefficients for a lower frequency, the wavelet is dillated by applying a scale factor $a$ with a large value, which results in a low-frequency $\Psi$. To isolate a time series of wavelet coefficients for a higher frequency, the wavelet is compressed by applying a scale factor $a$ with a small value, which results in a high-frequency $\Psi$.

According to Heisenberg’s uncertainty principle, there is a trade-off between the temporal precision and the spectral precision of the transform. Because the primary goal was to localize power fluctuations in the time domain, the bandwidth was deliberately chosen to be as narrow as possible to maximize the temporal precision of the transform, while maintaining at least two full cycles. The mother wavelet was compressed and applied at six scales, corresponding to frequencies of 5-30 Hz, in steps of 5 Hz. These frequencies were chosen for computational reasons. Namely, the time-frequency transform will multiply the total data size by a factor equal to the number of scales. At the same time, 5 Hz spacing roughly corresponds to the separation between classical EEG frequency bands and ensures that the spectrum is adequately sampled. The corresponding pseudo-frequencies ($F_{a}$) were estimated as the inverse of the product of the scale ($a$) and digitization interval ($\Delta$):

$$F_{a} = \frac{F_c}{a \cdot \Delta}. \quad (2.3)$$

### 2.2.3 Inter-event time distributions

Event times were identified by searching for all local maxima in the scalogram. Fig. 2.2 illustrates the procedure for detecting local extrema in a digitized signal. The original digital signal (top), representing the time series of real-valued coefficients obtained from the wavelet transform, is first converted to a difference function (bottom) by
Figure 2.1: **Wavelet transform.** (A) A wavelet transform is computed by convolving the wavelet with a section of the signal and shifting the wavelet through the whole signal. To isolate different frequencies, the wavelet is dilated or compressed and the procedure is repeated. (B) The wavelet transform can be used to isolate lower frequencies by dilating the wavelet, as well as higher frequencies, by compressing the wavelet.
subtracting adjacent data points from each other. For instance, the if the original signal is denoted as \( X \) with length \( m \), the difference function \( Y \) with length \((m - 1)\) is given by

\[
Y = \left[ X(2) - X(1) \quad X(3) - X(2) \quad \cdots \quad X(m) - X(m - 1) \right].
\]

The difference function approximates the first derivative for the original signal. Local extrema were detected and localized by finding the time points closest to where the difference function crossed zero (Fig. 2.2). Local maxima were sub-selected as those where the difference function had a negative slope while crossing zero, i.e. went from a positive to a negative value.

To prevent minute and insignificant peaks from being selected, a local neighborhood threshold was set as a ratio (5%) of the range of the scalogram amplitude. The exact choice of the ratio in the range 2-10% did not impact the functional form or the parameters of the inter-event time distributions in any significant manner.

The time between successive events (inter-event time) was calculated for each participant, condition, channel and wavelet scale (Fig. 2.3C), producing samples with an average of \( 904.2 \pm 50.0 \) inter-event times. Distributions of inter-event times were then fitted with the two-parameter gamma probability distribution function using maximum likelihood estimation (Fig. 2.3D). The two free parameters estimated were the shape \( k \) and scale \( \theta \). Fig. 2.4 shows how the look of the distribution is influenced by the two parameters. The gamma probability density has the following form:

\[
f(x) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}.
\]  

(2.4)

The gamma distribution was not selected \textit{a priori}, but was determined to be the most appropriate distribution when the data were fitted with 30 common distributions and the goodness of fit was assessed by way of the \( \chi^2 \) test using EasyFit software (MathWave Technologies). The fitted distributions were: Beta, Burr (Burr Type 12, or Singh-Maddala), Cauchy (Lorentz), Chi-Squared, Dagum (Burr Type 3, or Inverse Burr), Erlang, Error (Exponential Power, or Generalized Error), Error Function, Exponential, F Distribution, Fatigue Life (Birnbaum-Saunders), Frechet (Maximum Extreme Value Type 2), Gamma, Generalized Gamma, Gumbel Max (Maximum Extreme Value Type 1), Gumbel Min (Minimum Extreme Value Type 1), Hyperbolic Secant, Inverse Gaussian, Johnson SB, Johnson SU, Kumaraswamy, Levy, Laplace (Double Exponential), Logistic, Log-Gamma, Log-Logistic (Fisk), Lognormal, Nakagami (Nakagami-m),
Figure 2.2: **Peak picking.** Local maxima are extracted from the time series of wavelet coefficients (top) using the corresponding difference function (bottom). The difference function is computed by subtracting adjacent points from each other, approximating the first derivative. The points in time where the difference function changes from positive to negative values and in doing so crosses zero correspond to local maxima.

Normal (Gaussian), Pareto - first kind, Pareto - second kind (Lomax), Pearson Type 5 (Inverse Gamma), Pearson Type 6 (Beta dist. of the second kind), Pert, Power Function, Rayleigh, Reciprocal, Rice (Ricean, or Nakagami-n), Student’s t, Triangular, Uniform and Weibull.

The test statistic was significantly greater than the critical value for all 30 distributions, indicating significant departure from all those distributions. However, the gamma distribution had the lowest $\chi^2$ value across all fits and was ranked as the
Figure 2.3: **Overview of processing pipeline.** (A) The electroencephalogram is transformed into the time-frequency domain (B) using complex wavelets. (B) Local maxima in the scalogram (red vertical lines) are identified and marked as events. (C) The delay between successive events (inter-event time) is calculated in terms of digitization intervals. (D) The empirical probability distribution of inter-event times is fitted using the two-parameter gamma distribution function.

best-fitting distribution. Other common goodness of fit tests, such as the Kolmogorov-Smirnov and Anderson-Darling, were deemed inappropriate because they do not adjust the critical value to account for the degrees of freedom lost when parameters are estimated from the data. Upon visual inspection of the histograms it was clear that the two-parameter gamma distribution offered an excellent fit to the observed data (Fig. 2.5).

### 2.2.4 Gamma distribution

For completeness, I briefly review how the gamma distribution may arise as a sum of $k$ exponentially-distributed integrative stages. The time spent at the $i^{th}$ stage, $Y_i$, is
Figure 2.4: **Gamma distribution.** The two parameter gamma distribution is specified in terms of a shape parameter and a scale parameter.
Figure 2.5: **Empirical probability distributions.** Empirical probability distributions for the inter-departure time \( \tau \) for all subjects at one representative channel (Cz) and one representative frequency (15 Hz). Fitted gamma density functions are displayed in blue.

exponentially distributed with rate parameter \( k\mu \)

\[
h(y) = k\mu e^{-k\mu y}. \tag{2.5}
\]

The Laplace transform of \( Y_i \) is given by

\[
H^*(s) = \frac{k\mu}{s + k\mu}. \tag{2.6}
\]

By extension, the Laplace transform of the sum of \( k \) such random variables \((B^*)\) is the product of their individual transforms

\[
B^*(s) = \left( \frac{k\mu}{s + k\mu} \right)^k. \tag{2.7}
\]

The transform is then inverted to give the distribution of total service time \( b \)

\[
b(x) = k\mu (k\mu x)^{k-1} \frac{e^{-k\mu x}}{(k - 1)!}. \tag{2.8}
\]
which is a special case of the gamma distribution (Eq. 2.4) when $k$ is a positive integer and the scale parameter $\theta$ is the inverse of the exponential rate parameter ($k\mu = \theta^{-1}$).

A simple generative model for a gamma-distributed waiting time is the situation where several steps or stages must be completed between successive events and the waiting time at each of these stages is itself exponentially distributed (Fig. 2.6). The shape parameter ($k$) corresponds to the number of stages and the scale parameter ($\theta$) to their mean duration. What do gamma-distributed inter-event times reveal about brain function? Peaks in measured electromagnetic signal power arise due to interactions within and between neuronal ensembles. The time between two successive events in the scalogram may be thought of as a sum of shorter waiting times, during which afferent signals from multiple sources must be integrated.

Thus, the parameters of the gamma distributed inter-event times reveal how information is assimilated to give rise to the observed neural activity. Specifically, the time between successive events can be thought of as a sum of integrative steps and changes in either parameter due to experimental manipulation can be interpreted from this perspective. For instance, if inter-event time distributions recovered from neural activity during condition A have a greater shape parameter than during condition B, this would suggest that additional integrative steps are required to generate events during condition A. If distributions during task A have a greater scale parameter than during condition B, this would indicate that each of the integrative steps proceeds at a slower rate during condition A.

### 2.2.5 Partial least-squares (PLS) analysis

Each of the two parameters from the fitted gamma distributions ($k$ and $\theta$) were treated as measures of neural activity. A separate mean-centered partial least-squares (PLS) (McIntosh et al. 1996; Lobaugh et al. 2001; McIntosh and Lobaugh 2004; McIntosh and Mišić 2012) analysis was performed for each parameter. PLS analysis is a multivariate statistical technique that can be used to relate two “blocks” or sets of variables to each other (McIntosh et al. 1996; Lobaugh et al. 2001; McIntosh and Lobaugh 2004; McIntosh and Mišić 2012). In the context of neuroimaging, one set of variables may be exogenous, such as the study design (e.g. conditions), while the other may represent a set of endogenous variables, such as neural activity (e.g. $k$ or $\theta$) that varies across one or more dimensions (e.g. electrodes and frequencies). In the present study we
Figure 2.6: **Gamma as a sum of exponentials.** A gamma-distributed random variable with shape parameter $k$ and scale parameter $\theta$ may arise as a sum of $k$ exponentially-distributed random variables with mean equal to $\theta$. In queueing theory this corresponds to the situation where several stages of processing must be completed between successive events. The total time taken to complete all stages is gamma distributed.
related the differentiation between age groups to electrode- and frequency-dependent patterns of inter-event time distribution parameters.

In PLS this is achieved by computing the covariance matrix between the two sets of variables and decomposing this matrix into mutually-orthogonal “latent variables” using singular value decomposition (SVD) (Eckart and Young 1936). Each latent variable represents a particular relationship between the study design on one hand, and neural activity on the other. Specifically, each latent variable is expressed as a vector of design saliences and a vector of electrode saliences, as well as a scalar singular value (s). The elements of the design salience vectors are interpreted as a contrast between conditions, while the electrode saliences represent a particular pattern of electrodes and frequencies that maximally express that contrast. The singular value reflects the proportion of covariance between the design variables (groups and conditions) and electrophysiological variables (electrodes and frequencies) that is accounted for by each latent variable. This allows effect size to be estimated as the ratio of the square of the singular value associated with that particular latent variable to the sum of all squared singular values derived from the decomposition.

Nonparametric resampling techniques were used to assess the statistical significance and reliability of experimental effects. For each effect, statistical significance was determined using permutation tests. The rows (i.e., the observations) of the neural activity data matrix were randomly reordered (permuted) and the new data were subjected to SVD as before, to obtain a new set of singular values. These singular values are effectively generated under the null hypothesis that there is no association between neural activity and the task. The procedure was repeated 500 times to generate a sampling distribution of singular values under the null hypothesis. Since the singular value is proportional to the magnitude of the effect, a p-value is estimated as the probability that singular values from the distribution of permuted samples exceed the singular value from the original, non-permuted data.

To determine which electrode saliences were reliable, their standard errors were estimated using bootstrap resampling (Efron and Tibshirani 1986). Bootstrap samples were generated by random sampling with replacement of participants within conditions (500 replications). Saliences were deemed to be reliable if the 99% confidence interval did not include zero. Under the assumption that the bootstrap distribution is unit normal, this condition holds if and only if the absolute value of the ratio of the salience to its bootstrap-estimated standard error is greater than or equal to 2.57 (Efron
and Tibshirani 1986). Bootstrap ratios were thresholded across all data points to allow parsimonious identification of latent variables.

### 2.3 Results

#### 2.3.1 Inter-event time distributions

The empirical inter-event time distributions were fitted with the two-parameter gamma distribution for each condition, subject, electrode and frequency. The gamma distribution offered a good fit at all frequencies. Despite some individual differences in the parameters of the distribution, the form was remarkably consistent across subjects. This is illustrated in Fig. 2.5, which shows the fits for all 56 subjects at one electrode and one frequency. Nevertheless, there was also substantial variation from subject to subject for both estimated parameters. To illustrate the individual variation of fits across frequencies, we also report the coefficient of variation of each parameter in the Eyes-Open condition, for electrode Cz, for the six frequencies, going from 5 to 30 Hz: 0.22, 0.20, 0.21, 0.20, 0.25 for the shape parameter; 0.33, 0.32, 0.34, 0.34, 0.37 and 0.31 for the scale parameter. The data indicate that both parameters are quite sensitive to individual differences. The spatial distributions of group means for $k$ and $\theta$ are displayed in Figs. 2.7 and 2.8 and discussed in more detail in the following subsection.

![Figure 2.7: Group means for the shape parameter.](image)

**Figure 2.7:** Group means for the shape parameter. Group means are displayed separately for eyes-closed (top row) and eyes-open (bottom row) conditions, and frequencies from 5 to 30 Hz.
Figure 2.8: **Group means for the shape parameter.** Group means are displayed separately for eyes-closed (top row) and eyes-open (bottom row) conditions, and frequencies from 5 to 30 Hz.

Note that since wavelets effectively act as a band-pass filter, the means of inter-event time distributions should vary in proportion to frequency, such that departures are expected to occur at a faster rate at higher frequencies, resulting in shorter mean inter-event times. As an example, the group mean inter-event times for the Eyes-Open condition, channel 60, going from 5 Hz to 30 Hz, were $50.9 \pm 1.3, 47.9 \pm 1.2, 45.3 \pm 1.2, 44.3 \pm 1.1, 43.1 \pm 1.1$ and $40.9 \pm 0.8$ ms. However, these analyses were concerned with identifying regional and state-dependent statistical effects and did not compare frequencies to each other.

### 2.3.2 Shape parameter $k$

Across all frequencies, the shape parameter of the fitted gamma distributions was greater over the posterior (occipital and parietal) channels (Fig. 2.7). Moreover, this measure was sensitive to experimental condition and was greater in the eyes-closed than in the eyes-open condition (Fig. 2.7), an observation statistically supported by the PLS analysis ($p \ll 0.001, s = 11.69$). The statistical effect was most reliable across all frequency bands over occipital channels and to a lesser extent over parietal and frontal channels (Fig. 2.9, top row).
Figure 2.9: Bootstrap ratio maps for two PLS analyses comparing eyes-closed and eyes-open conditions. Top row: the dependent variable in the first analysis is the shape parameter $k$ and bootstrap ratios indicate channels at which values are greater in the eyes-closed versus the eyes-open condition. Bottom row: the dependent variable in the second analysis is the scale parameter $\theta$ and bootstrap ratios indicate channels at which values are greater in the eyes-open versus the eyes-closed condition.

2.3.3 Scale parameter $\theta$

The scale parameter was lower at most posterior and vertical channels and generally much higher over temporal and anterior channels. This pattern was observed at all frequencies (Fig. 2.8). Values were significantly greater in the eyes-open condition ($p \ll 0.001, s = 86.02$) and this effect was most stable over occipital channels (Fig. 2.9, bottom row). There was also some suggestion of frequency dependence in the sense that the bootstrap ratios were slightly higher (i.e. the effect was more robust) at lower frequencies. It is worth noting that the most extreme values of $\theta$ were observed at electrodes close to the eyes (Fig. 2.8), which tend to undergo the heaviest signal processing under most artifact rejection schemes. However, this did not appear to affect the statistical analysis, as the condition differences at these electrodes were not reliable by bootstrap test.
2.4 Discussion

The present study describes a signal processing method that can be used to identify putative message departure times from neurophysiological data and quantify the distribution of times between successive departures. The analysis demonstrates that the two-parameter gamma distribution offers a good fit to the inter-event time distribution. The parameters of inter-event time distributions were not uniform across the scalp and instead displayed spatial specificity. Namely, distributions recovered from medial posterior electrodes had larger $k$ and smaller $\theta$ compared to anterior electrodes. This suggests that inter-event times may be sensitive to regional differences in connectivity and/or processing capacity. In addition, as “proof of concept”, inter-event times are shown to be sensitive to the presence of visual input, with significantly greater $\theta$ and smaller $k$ at occipital channels when participants kept their eyes open.

Variability of inter-event times

What does systematic variation in $k$ and $\theta$ tell us about the functional capacity of the underlying system? For example, what does it actually mean for a cortical region to produce inter-event times with greater $\theta$ and smaller $k$ in the eyes-open condition? Here it may be instructive to consider other statistics of the distribution that are easier to interpret. For example, the coefficient of variation ($c_V$, the ratio of the standard deviation to the mean) is a normalized measure of dispersion and for the gamma distribution is given by

$$c_V = \frac{1}{\sqrt{k}} \quad (2.9)$$

Thus, inter-event times were more variable at medial posterior channels compared with the rest of the scalp. Moreover, the distributions became more dispersed in the eyes-open condition and the effect was robust at occipital channels. These results suggest that inter-event times capture a facet of network traffic. For example, traffic traces in telecommunication networks are found to be more variable under conditions of greater spectrum occupancy (Riihijärvi et al. 2008, 2009). The fact that inter-event times were more variable at parietal channels is consistent with the notion that structures situated in posterior cortex (particularly close to the midline, such as the precuneus and posterior cingulate) enjoy an exalted status in the connectome. These regions tend to occupy positions along the shortest white-matter paths between all

Given that the biggest difference between eyes-open and eyes-closed is the availability of visual input it is not surprising that condition differences were expressed most reliably over the occipital portion of the scalp. This condition-dependent differentiation may reflect the transient reconfiguration of functional networks in response to changes in external input. For instance, as visual processing becomes more prominent in the eyes-open condition, more information should be routed through the occipital cortices. This should influence the rate of information exchange and total flux through the associated subnetworks, making the underlying biological and cognitive operations less regular and less predictable. This is reflected by the results, which indicate that when the eyes are open, both very short and very long inter-event times become more likely than when the eyes are closed. The expression of condition differences at multiple frequencies precludes the interpretation that they are the result of a simple difference in power spectral density in the \( \alpha \) frequency band typically observed in visual tasks. For example, condition differences were not specific to activity resolved at 10 and 15 Hz.

### 2.4.1 Importance of the gamma distribution

From the perspective of telecommunication systems, the fact that inter-event times were best approximated by the gamma distribution is significant. The gamma distribution arises naturally and often in such systems, particularly in relation to waiting times (Mukherjee 1994). In particular, when the shape parameter \( k \) is a positive integer, the gamma distribution can be thought of as the sum of \( k \) independent exponentially distributed random variables, each with a rate parameter \( k \mu \). This situation arises when a message must be processed or receive some type of service over a series of stations or stages at a server (termed an Erlang server, Fig. 2.6), each of which has an exponential service time distribution. For instance, the server may represent a population of neurons (as in the graph model). The stages are simply a sequence of processes that take place before a unit of information is emitted. In the context of a neuronal ensemble, these processes may represent the interactions among cells within
the ensemble. The time spent at the \(i^{th}\) stage, \(Y_i\), is drawn from the probability density function

\[
h(y) = k\mu e^{-k\mu y}.
\]  

(2.10)

Since the service times are exponential, the expectation and variance for \(Y_i\) are given by:

\[
E[Y_i] = \frac{1}{k\mu},
\]  

(2.11)

\[
Var[Y_i] = \left(\frac{1}{k\mu}\right)^2.
\]  

(2.12)

The total time spent at the server (traversing the \(k\) stages) is the sum of \(k\) independent identically distributed random variables drawn from the distribution \(h\). Therefore, the expectation and variance of the total processing time \(Y = \sum_{i=1}^{k} Y_i\) can be calculated by summing across the \(k\) stages:

\[
E[Y] = k\left(\frac{1}{k\mu}\right) = \frac{1}{\mu},
\]  

(2.13)

\[
Var[Y] = k\left(\frac{1}{k\mu}\right)^2 = \frac{1}{k\mu^2}.
\]  

(2.14)

Importantly, the coefficient of variation of the total service time is given by

\[
c_v = \frac{1}{\sqrt{k}},
\]

resulting in a hypoexponential service time distribution, named to denote the fact that the coefficient of variation for this distribution is smaller than that of the exponential distribution (i.e. 1) (Kleinrock 1976a). Hypoexponential service times indicate that the underlying processing stages are arranged in series (Fig. 2.6). If there is any branching and some stations are arranged in parallel, service time distributions will be hyperexponential, with a coefficient of variation greater than 1 (for a detailed derivation see (Kleinrock 1976a)). In the present data, inter-event times were found to be hypoexponential, which under this theoretical framework is indicative of the former arrangement. This view is biologically plausible, because it suggests that once a unit of information arrives to a node, the sequence of operations performed on that unit is set and does not change from unit to unit. Note however, that although these stages may represent a transformative process, they do not necessarily alter the information content of each unit. Importantly, this derivation should not be misinterpreted as a statement about whether large-scale cognitive processes are coordinated in series or in
parallel. The results merely suggest that there is no variation in the sequence of steps performed on each unit.

Overall, this conceptualization of neural dynamics provides a novel narrative of information flow in the brain. This view suggests that units of information may be processed in a series of independent stages. Moreover, the number of processing stages \((k)\) and duration of each stage \((\theta)\) vary across regions of the brain and depend on internal and external conditions. The presence of visual input appears to engender a mode of operation with fewer processing stages but slower service rates. Thus, although it is not the only possible explanation, a telecommunication-based perspective offers a simple and biologically meaningful interpretation for the observed hypoexponential gamma-distributed inter-event times and the associated parameters \(k\) and \(\theta\).

### 2.4.2 Methodological considerations

The idea to delineate signal units in the EEG and to characterize the sequence of inter-event times is directly inspired by research in telecommunication networks. However, it is important to consider the physiological validity of the telecommunication model. To what degree are units of information recovered from the EEG scalogram comparable to data transmitted in a typical telecommunication network? In the present approach, local extrema in the scalogram are de facto the basic units of information transfer, whereas in neural systems the more likely candidates would be action potential spikes or spike trains (Dayan and Abbott 2001). Thus, the ideal analysis of the “output process” at individual neurons would be based on single-cell recordings of spiking activity. The decomposition of EEG recordings used in the present study approximates such an analysis.

The key is that we would like to know how information emitted across the scalp changes under different experimental conditions. For this context and by virtue of their spatial scale and coverage, gross neurophysiological recordings such as the EEG, which represent aggregated postsynaptic potentials from entire populations of neurons are the more appropriate measure of neural activity from which to isolate inter-event times compared to single cell recordings. It is also interesting to note that, although action potential spikes are typically modeled as a Poisson point process, inter-spike intervals (ISIs) measured from single cells often do not appear exponential but take on a functional form rather more similar to the gamma distribution described here (e.g.
Fig. 1C in (Shadlen and Newsome 1998)).

How will the present method generalize to other experimental settings, such as an event-related design with multiple shorter trials? One of the keys to fitting distributions to empirical data is sufficient sample size. In other words, to estimate the distribution of packet inter-event times with a reasonable degree of confidence, one must generate many such packet departures. In a more traditional setting where time series are epoched into shorter segments the same procedure could be applied by calculating inter-event times in all individual trials and collating them into a single sample to be fitted. In addition, it remains unclear what impact, if any, time-locked evoked responses would have on inter-event times and this certainly warrants further investigation.

The EEG is vulnerable to volume conduction and therefore the spatial precision with which we were able to describe changes in inter-event time distributions is naturally limited. Moreover, the present method treats all units of information in the same vein, even though peaks in the EEG scalogram vary in their amplitude. In other words, our method implicitly allows the possibility that units of information transmitted in the brain may vary in size. However, even if differences in message size were to be taken into account, this would not change the inter-event time statistics extracted from the time series.

A possible extension to the present work would be to investigate whether this method could be used to recover the shape and scale parameters from a process that had been generated using the Erlang model. The biggest methodological challenge would be to specify an appropriate forward model that can translate the stochastic process generated by the Erlang model to electrical activity. The importance of this approach would be to demonstrate that the wavelet method presented here could be used to accurately recover the underlying scale and shape parameters.

2.4.3 Conclusion

In this study I applied tools from teletraffic engineering to the study of neural dynamics. I have developed a way to identify electrophysiological events that may be interpreted as departing units of information and have shown that the times between departures are distributed according to the gamma probability distribution. In addition, I have demonstrated that this facet of neural activity is meaningful from the perspective of
cognitive function. Namely, distributions of inter-event times are highly dependent on cognitive state and spatial location. I conjecture that inter-event times reflect the flow of network traffic and index the communication capability of the brain’s functional architecture.
Chapter 3

Inter-event times and face processing

adapted from Mišić et al. (in press) J Cogn Neurosci

3.1 Introduction

The brain is a complex network of distributed regions that must integrate and communicate information to engender perception, cognition and motor function (Stam and van Straaten 2012). Thus, it is important to understand how streams of information are assimilated and subsequently emitted from individual nodes. The rate at which nodes receive, process and output information are potentially informative aspects of the structural and functional architecture of brain networks.

To this end, a rich repertoire of techniques have emerged to capture the flow and exchange of information in brain networks. For instance, connectivity-based methods can quantify the topological role of brain regions in the context of network communication (Sporns et al. 2000; Sporns and Zwi 2004; Stam et al. 2007; Hagmann et al. 2008; Rubinov et al. 2009a; van den Heuvel and Sporns 2011; Mišić et al. 2011b). Neural activity may also be treated as an expression of a dynamical system in order to characterize various aspects of information processing. Entropy-based metrics can be used to estimate the rate at which information is generated by that system (Stam 2005; McIntosh et al. 2008; Mišić et al. 2010; Vakorin et al. 2011a,b). Likewise, the dynamic range of the system may be studied in terms of the variability of neural activity (Garrett et al. 2010, 2011, 2012). Finally, time series of neural activity can be segregated into a set of states, in order to study transitions among those states (Freyer et al. 2009, 2012; Pascual-Marqui et al. 1995). Yet despite considerable progress, we lack complete
understanding of how information is integrated and projected to other nodes in the network.

The role of integration in the context of networked computation is of interest not only in neuroscience, but in network science in general. In the previous chapter I demonstrated that tools from teletraffic engineering could be adapted to study information flow in brain networks (Mišić et al. 2011a). A fundamental statistic for communication networks is the distribution of time between successive message departures at each node, because it reflects the nature of the arrivals to the node as well as how the node processes these arrivals. I showed that the resting-state electroencephalogram could be partitioned into a trace of output events, analogous to a departure of a unit of information.

However, information is dynamically routed through brain networks over multiple time scales. Long-term changes such as pruning and myelination that take place during development transform the configuration of anatomical brain networks (Hagmann et al. 2010) and the paths available for information transfer. Likewise, transient interareal functional associations may form spontaneously or in response to external demand, which suggests that information flow can be flexibly re-directed (Stephan et al. 2003; McIntosh et al. 1994; Honey et al. 2009). The purpose of this study is to investigate how the reconfiguration of structural and functional networks affects the way in which information is integrated. Specifically, I sought to determine whether development and cognitive engagement affect the functional form of inter-event time distributions (indicating an entirely different mode of functioning) or just the parameters of those distributions (indicating an alteration within the same mode).

To study regionally-specific dynamics, neural activity was estimated at several hundred sources (Fig. 3.1) in the brain using a beamforming source analysis on magnetoencephalography (MEG) data recorded while children and adults performed a face recognition task with upright and inverted faces. I hypothesized that developmental changes in inter-event times should be observed throughout the brain, because in the age range of this sample most areas experience structural change due to myelination and synaptic pruning (Lenroot and Giedd 2006; Paus et al. 1999; Taylor 2006).

A face recognition task allowed us to characterize the integration and subsequent output of neural processes during a cognitive task. The face inversion effect was investigated, whereby recognition of faces is impaired if they are presented upside-down. A similar effect is observed for other types of visual stimuli such as everyday
objects, but the impairment is disproportionately large for faces (Yin 1969). Face inversion is thought to disrupt the configural (Rhodes et al. 1993; Freire and Lee 2001; Mondloch et al. 2002) and/or holistic (Tanaka and Farah 1993; Farah et al. 1995) information about a face, leading to significant behavioural and neural differences in processing of upright and inverted faces, in adults and children (Taylor et al. 2004). I hypothesized that if integration of information about the relations among features as well as their overall arrangement is affected by face inversion, inter-event time distributions should be different for upright and inverted faces.

Output events were defined as the local maxima of the MEG scalogram (Fig. 3.2), which follows from the physiological interpretation of gross electromagnetic brain signals such as EEG and MEG. Namely, peaks in electromagnetic signal power represent synchronous post-synaptic potentials from a population of neurons. Given that neuron soma comprise a large component of grey matter nodes of brain networks, then post-synaptic potentials may be thought of as an output event. Inter-event times were fitted with the two-parameter gamma distribution. The goal of this study was to identify spatial patterns of task-dependent changes in inter-event times that were expressed across multiple age groups, without an a priori region of interest. Thus, a multivariate analytic approach (partial least squares) (McIntosh and Lobaugh 2004) was used, because it enabled group and task differences to be statistically assessed simultaneously across all sources and temporal scales.

Changes in the inter-event distributions are interpreted in terms of a particular generative model - the Erlang server - which models how information accumulates at a particular node in a network. Note however, that finding a Gamma distribution does not imply this particular model and in fact, the underlying process cannot be inferred directly from the functional form of the distribution. Therefore, the interpretation of these results in terms of the Erlang model is strictly speculative.

### 3.2 Materials and methods

#### 3.2.1 Participants

Seventy-one naïve, healthy participants took part in the study, including 49 children, aged 6-7 ($n=10$, 3 female), 8-9 ($n=10$, 9 female), 10-11 ($n=9$, 2 female), 12-13 ($n=10$, 6 female), 14-16 ($n=10$, 3 female) and 22 adults, aged 20-41 years (mean age 25.7 years,
9 female). None of the participants wore any metallic implants or had ferromagnetic dental work. All participants reported normal or corrected-to-normal vision. Experiments were performed with the approval of the Research Ethics Board at the Hospital for Sick Children.

3.2.2 Stimuli and task

Participants performed a simple 1-back task in which they were asked to judge whether stimulus presented at the time was the same or different compared to the previous one. The stimulus set consisted of 240 greyscale photographs of faces of young adults (2.4 x 3° visual angle) with neutral expressions that were unfamiliar to the participants. All faces were without any paraphernalia or facial hair. Male and female faces were equiprobable. In each block of 180 trials, one third of the faces immediately repeated, such that in each block 120 faces were new and 60 were repeated. To avoid any confounding effect of memory, only the trials that contained new presentations of faces were analyzed. Upright faces were presented in one block and inverted faces in the other, with the order of the two blocks counterbalanced across participants.

Stimuli were back-projected via two mirrors to a screen positioned at a viewing distance of 70 cm from the participant. Faces were presented centrally over a uniform black background for 500 ms. The duration of the inter-stimulus interval varied randomly and with equal probability between 1200 and 1500 ms and during this time a white fixation cross was displayed at the center of the screen. Subjects were instructed to respond as quickly and accurately as possible to indicate if a face had repeated. Stimulus presentation and responses were controlled using Presentation software (Neurobehavioral Systems Inc.).

3.2.3 Magnetoencephalogram (MEG) acquisition

MEG data were recorded using a 151-channel whole-head CTF system (MEG International Services Ltd.). Subjects lay supine in a dimly-lit magnetically shielded room at the Hospital for Sick Children in Toronto. Prior to acquisition, three localization coils were placed at the nasion and bilateral preauricular points to help localize the participant’s head relative to the MEG sensor array at the start and finish of each block. Motion tolerance was limited to 1 cm and neuromagnetic activity was digitized at a
Event-related beamforming was used to estimate neural activity at 529 sources evenly distributed across the brain.

Data were filtered off-line with a band-pass of [0.5 100] Hz. Data were then epoched into [-100 1500] ms segments time-locked to stimulus onset. Following the MEG recording session, the three fiducial coils were replaced by MRI-visible contrast markers and 3-D SPGR (T1-weighted) anatomical images were acquired using a 1.5 T Signa Advantage system (GE Medical Systems).

3.2.4 Event-related beamforming (ERB)

ERB is a 3D adaptive spatial filter that uses surface field measurements to estimate activity at desired locations in the brain (Robinson and Vrba 1999; Sekihara et al. 2005; Cheyne et al. 2007). Individual anatomical MR images were warped into the standard Talairach space using a nonlinear transform in SPM2. 529 source locations were chosen on a grid of size 5 mm such that they were uniformly spaced and sufficiently few in number to allow reasonable computation time (Fig. 3.1). These locations were then warped back into the individual participants’ brains using the inverse transform. Activity at each target source was estimated as a weighted sum of the surface field measurements (Osterman et al. 1959). Weight parameters and the orientation of the source dipole were optimized in the least squares sense, such that the average power originating from all other locations was maximally attenuated without any change to the power of the forward solution associated with the target source. The forward solution for the beamformer was modeled by fitting multiple sphere models to the inner skull surface of each participant’s MRI using BrainSuite software (Shattuck and Leahy 2002). The weights were then used to compute single-trial time series for each source.
3.2.5 Wavelet transform

The wavelet transform, event identification and statistical assessment were performed in a manner analogous to Chapter 2. The present chapter includes only essential details; for more information, please consult the Methods section in Chapter 2.

Instantaneous changes in signal power were studied in the time-frequency domain by applying a wavelet transform (Torrence and Compo 1998). Individual epochs were convolved with a Morlet wavelet and signal power was estimated as the modulus squared of the real part of the wavelet coefficients. The mother wavelet had center frequency \( F_c \) equal to 1 Hz and bandwidth (standard deviation) of the Gaussian envelope equal to 2 s. Wavelet parameters were chosen to maximize the temporal precision of the transform. Thus, the bandwidth was made as narrow as possible while maintaining at least two full cycles of the wavelet. The mother wavelet was compressed and applied in sliding windows at six different scales, corresponding to frequencies of 5-30 Hz, in steps of 5 Hz.

3.2.6 Inter-event time distributions

The output process at each source was defined in terms of the peaks and troughs in the scalograms (Fig. 3.2). Event times were identified by searching for all local maxima in the scalogram. To prevent insignificant peaks from being selected, a local neighborhood threshold was set as a ratio (5%) of the range of the scalogram amplitude. The exact choice of threshold did not impact the functional form of the distributions in any significant manner, such that thresholds varying from 2 to 10% gave rise to very similar distributions. The times between successive events were calculated for each participant, condition, source and wavelet scale.

The statistical properties of the inter-event times at each source were characterized by modeling the functional form of the distribution (Fig. 3.2). EasyFit software (MathWave Technologies) was used to fit the inter-event time data with 30 common distributions by maximum likelihood estimation. The complete list of tested distributions is included in Chapter 2. The distributions were ranked according to the \( \chi^2 \) statistic, with the two-parameter gamma distribution offering the best fit across all conditions, participants, sources and frequencies.

As in the previous chapter, the Erlang server is used as a simple generative model for the gamma-distributed stochastic process. The model represents a situation where
Figure 3.2: **Processing pipeline.** The source activity for each individual trial was convolved with a complex Morlet wavelet to estimate frequency- and time-dependent fluctuations in power. Supra-threshold troughs in the scalogram were marked as events (red vertical lines) and the time between successive events was calculated. A histogram of all inter-event times was constructed by collating the inter-event times from all trials in a given condition. Maximum likelihood estimation was used to fit a gamma probability density function to the sample.
several steps or stages must be completed between successive events and the waiting time at each of these stages is itself exponentially distributed. The shape parameter ($k$) corresponds to the number of stages and the scale parameter ($\theta$) to their mean duration.

Thus, the time between successive events can be thought of as a sum of integrative steps and changes in either parameter due to experimental manipulation can be interpreted from this perspective. For instance, if inter-event time distributions recovered from neural activity during task A have a greater shape parameter than during task B, this would suggest that additional integrative steps are required to generate events during task A. If distributions during task A have a greater scale parameter than during task B, this would indicate that each of the integrative steps proceeds at a slower rate during task A.

Note however, that this is not the only possible model that can generate gamma statistics. Although queueing network models are often used in cognitive psychology (Liu 1996), diffusion models are also commonly used to explain gamma distributions (Ratcliff 1979; Van Zandt and Ratcliff 1995). However, there is no way to objectively assess which model is most suitable. In the present study, we chose to interpret changes in gamma statistics from the perspective of the Erlang model because of its simplicity and because it provides a taxonomy that fits well with the well-established notions of configural and holistic processing of faces.

### 3.2.7 Partial least-squares (PLS) analysis

The two fitted gamma distribution parameters $k$ and $\theta$ were treated as dependent measures of neural activity and mean-centering PLS analysis was used to assess the differences between age groups and conditions for each of these measures. PLS is described in detail in the previous chapter, and is applied in an analogous way here. Namely, the two “blocks” or sets of variables that we sought to relate were the study design (age groups and conditions) and neural activity ($k$ and $\theta$) that varied across sources and frequencies. Statistical maps were generated using tessellation-based linear interpolation to estimate bootstrap ratios for each voxel. Bootstrap ratios were thresholded across all voxels to allow parsimonious identification of latent variables.
Figure 3.3: **Individual inter-event time distributions.** Inter-event time distributions are displayed for five randomly-selected subjects in each of the six age groups, at one representative source and one representative frequency. Each of the distributions has also been fitted with a two-parameter gamma probability density (blue lines).

### 3.3 Results

#### 3.3.1 Inter-event time distributions

Inter-event time distributions were fitted with thirty different probability distributions, separately for each participant, condition, source and frequency. The first finding is that the functional form of the distributions was remarkably consistent across participants and that the two-parameter gamma distribution offered the best fit out of all distributions tested. Importantly, the gamma distribution proved to be a good fit at all developmental time points, indicating that the functional form of the inter-event time distribution changed little during healthy brain development. This is illustrated in Fig. 3.3, where the inter-event time histograms and corresponding fits are displayed for five randomly-selected participants in each age group, at one representative source and frequency.

However, the shape and scale of the distributions were not identical for all participants and the parameters appeared to be sensitive both to individual differences and
Chapter 3. Inter-event times and face processing 47

maturation (Fig. 3.3). In the following sections I investigate whether the parameters of the inter-event time distributions systematically varied due to development and/or changes in cognitive processing due to face inversion. These changes are interpreted in terms of the Erlang model.

3.3.2 Development

To determine whether brain development was associated with systematic changes in the parameters of inter-event time distributions, we contrasted the six age groups to each other, separately for each condition. During maturation, the scale parameter $\theta$ increased at low frequencies and decreased at high frequencies, indicating slower and faster processing stages, respectively. The effect was observed for both inverted ($p \ll 0.001$, singular value, $s = 78.66$) and upright ($p \ll 0.001$, singular value, $s = 71.15$) faces, accounting for 63.45% and 58.71% of the covariance between age groups and $\theta$. This is evident in Fig. 3.4, where the ratios of the source saliences to their bootstrap-estimated SEs are generally positive at low frequencies and negative at high frequencies. The spatial distributions of these statistical trends at 5 and 15 Hz are shown in Fig. 3.4. Although the age-dependent increase at low frequencies was particularly reliable in the occipital lobe, the corresponding decrease at higher frequencies was more diffuse and could be observed in most brain regions.

With increasing age the shape parameter $k$ decreased at lower frequencies and increased at higher frequencies, indicating fewer integrative steps at lower frequencies and more integrative steps at higher frequencies. The effect can be observed in the frequency-wise transition from negative to positive bootstrap ratios at 10 Hz (Fig. 3.5). This trend was statistically significant for both inverted ($p \ll 0.001, s = 7.62$) and upright ($p \ll 0.001, s = 7.23$) faces and accounted for 75.58% and 72.96% of the cross-block covariance in the respective runs. As with the scale parameter, the age-driven decrease in the shape parameter at low frequencies was relatively restricted to the occipital lobe, while the increase at higher frequencies was robust throughout the brain (Fig. 3.5).
Figure 3.4: **Developmental effect captured by the scale parameter.** (A) A statistical contrast depicting the age-driven changes in the scale parameter. (B) Imagescale of the bootstrap ratios associated with the latent variable, showing which combinations of sources and frequencies reliably expressed the associated contrast. Together with the design scores, this indicates that during development the scale parameter increased at low frequencies and decreased at high frequencies. (C) and (D) Montage showing the bootstrap ratios interpolated for the whole brain at 5 Hz and 15 Hz.
Figure 3.5: Developmental effect captured by the shape parameter. (A) A statistical contrast depicting the age-driven changes in the shape parameter. (B) Imagescale of the bootstrap ratios associated with the latent variable, showing which combinations of sources and frequencies reliably expressed the associated contrast. Together with the design scores, this indicates that during development the shape parameter decreased at low frequencies and increased at high frequencies. (C) and (D) Montage showing the bootstrap ratios interpolated for the whole brain at 5 Hz and 15 Hz.
3.3.3 Development and face inversion

To investigate the effect of development and face processing simultaneously, runs with upright and inverted faces were contrasted across all age groups. In the youngest group the scale parameter $\theta$ was greater for inverted compared to upright faces, but with maturation this difference gradually became smaller and eventually reversed, such that in adulthood $\theta$ was greater for upright faces compared to inverted faces ($p \ll 0.001, s = 20.21$, corresponding to 25.0% of cross-block covariance) (Fig. 3.6). This effect was primarily positively expressed in right inferior temporal cortex at 15 Hz and negatively expressed in left inferior temporal cortex at 5 Hz. Although there was some contribution from the superior aspect of the right cerebellum, this is likely to reflect spatial smearing or leakage from inferior temporal sources due to the data-driven MEG source localization (Itier et al. 2006; Taylor et al. 2008).

The shape parameter $k$ of the fitted gamma distribution displayed a similar trend, in the sense that $k$ was greater for upright compared to inverted faces in the youngest group, but during development this inversion effect gradually reversed so that by late adolescence inverted faces were associated with greater $k$ than upright faces ($p \ll 0.001, s = 1.79$, corresponding to 23.38% of cross-block covariance) (Fig. 3.7). This trend was reliable mainly in the anterior portion of the right parahippocampal gyrus at 25 Hz. Note however that the spatiotemporal pattern associated with this effect was only able to capture face inversion effects in the non-adult groups. In the adult group a different spatiotemporal pattern differentiated upright and inverted faces ($p = 0.08, s = 1.01$, corresponding to 9.02% of cross-block covariance), mainly in the left inferior frontal cortex at 10 Hz (Fig. 3.7).

3.4 Discussion

3.4.1 Inter-event time distributions

The empirical probability distribution of inter-event times most closely resembled a two-parameter gamma distribution, in concordance with previous findings (Mišić et al. 2011a). The fact that the distribution translated from resting-state EEG recordings to MEG recordings with multiple shorter trials is encouraging because they measure the same biophysical process and should display similar dynamics. Interestingly, the func-
Figure 3.6: **Face inversion effect captured by the scale parameter.** Top: A statistical contrast depicting the difference between Upright and Inverted faces across age groups. Middle and bottom: Statistical maps of bootstrap ratios show brain regions that most reliably expressed the associated statistical contrast. Positive and negative bootstrap ratios are indicated by warm and cold colours respectively. Maps are thresholded at a bootstrap ratio value of $|2.57|$.
Figure 3.7: **Face inversion effect captured by the shape parameter.** The two mutually-orthogonal latent variables shown captured the greatest (A) and second-greatest (B) proportion of cross-block covariance in the same PLS analysis. Top: A statistical contrast depicting the difference between Upright and Inverted faces across age groups. Bottom: Statistical maps of bootstrap ratios show brain regions that most reliably expressed the associated statistical contrast. Positive and negative bootstrap ratios are indicated by warm and cold colours respectively. Maps are thresholded at a bootstrap ratio value of $|2.57|$.
tional form of the distribution was the same across groups and tasks, which suggests that even under changing network conditions the assimilative dynamics maintain a characteristic mode of operation. Experimental manipulation caused systematic variation in the parameters of the distribution, across regions and individuals, which we discuss in more detail in subsequent sections.

Our analyses reveal that inter-event times in human MEG display robust gamma statistics. In the following discussion, we offer an interpretation of experimentally-induced non-stationarities in terms of the Erlang model and information accumulation. We stress that this is merely a model and that our data do not prove the existence of such a mechanism. The main advantage of the Erlang model is that it allows us to relate changes in gamma statistics to holistic and configural processing in a simple and natural way.

### 3.4.2 Development

Neural development was characterized by changes in both the shape and scale of the inter-event distributions. At low frequencies (<10 Hz), increasing age was marked by fewer and slower integrative steps. At higher frequencies (≥10 Hz) the opposite was true, with increasing age marked by more integrative steps that were also faster. The differential effects observed for low and high frequencies are reminiscent of previous literature that shows that spectrum occupancy shifts during development, with decreased power in low frequencies and increased power in high frequencies (Gasser et al. 1988). Note that the scale and shape of a distribution are theoretically independent and the fact that one changes during development does not imply that the other will change as well. Therefore, the dissociation between low and high frequencies on one hand and scale and shape on the other suggests a gradual redistribution of communication over specific bands.

Interestingly, the low-frequency effects appeared to be more regionally specific than the high-frequency effects. In the lower frequency regime statistical trends were most prominent in occipital cortex, with additional contribution from superior paracentral regions in the case of the scale parameter. The fact that this particular trajectory was constrained mainly to areas with visual and motor utility suggests reconfiguration constrained by functional neuronatomy. This is not surprising, given that the stimuli used in the experimental protocol were faces, which are likely to engage the ventral
visual processing stream, and that participants were required to respond manually to these stimuli. Conversely, in the higher frequency regime statistical trends were spatially more broad and could be observed throughout the brain.

3.4.3 Development and face inversion

Inter-event times were also sensitive to face inversion: the statistical analysis revealed an interaction between development and face processing, with the inversion effect becoming more pronounced with increasing age. In addition, the inversion effects identified for both parameters were reversed in the youngest groups compared to the adolescent and adult groups. This type of interaction is also observed with the latency and amplitude of the face-sensitive N170 evoked potential and the M170 evoked field. For example, the adult N170 has both a greater (more negative) amplitude and a longer-latency peak in response to inverted faces (Bentin et al. 1996; Rossion et al. 1999, 2000; Itier and Taylor 2002; Taylor et al. 2011a). In young children the effect is often reversed, with greater amplitudes and longer latencies in response to upright faces (Taylor et al. 2001, 2004).

Task-dependent developmental trajectories were driven mainly by parts of inferior temporal cortex. From the perspective of the Erlang model, maturation was associated with progressively faster assimilation for upright compared to inverted faces in left inferior temporal cortex and slower assimilation in right inferior temporal cortex. Face perception has consistently been linked with inferior temporal cortex in general and fusiform gyrus in particular. For example, functional MRI (fMRI) studies frequently report increased activation in fusiform gyrus in response to faces compared to other complex visual stimuli (Puce et al. 1995; Kanwisher et al. 1997; McCarthy et al. 1997; Halgren et al. 1999; Haxby et al. 1999; Schiltz and Rossion 2006). Face inversion effects have also been reported, whereby upright faces elicit greater activation compared to inverted faces (Kanwisher et al. 1998; Haxby et al. 1999; Yovel and Kanwisher 2005). Increasing lateralization of brain functions with age is well documented and ERP studies have shown less right-sided dominance for face processing in children (Taylor et al. 1999, 2001). Face processing skills improve throughout childhood and into adulthood (Taylor et al. 2004) (but see also (Crookes and McKone 2009), so the opposing trends in right and left fusiform gyri may indicate that the right fusiform gyrus becomes increasingly specialized for faces.
The possibility that information about upright faces may be integrated in a slower, more deliberate manner in the right fusiform gyrus is consistent with the notion that faces are perceived in terms of the configuration and relative positions of facial features (configural processing), as well as a coherent whole (holistic processing). Both accounts suggest that face perception depends on intensified inter-areal integration to synthesize a complex percept and our data support this premise, because we find that face perception involves longer integrative steps, as measured by the scale parameter. Similarly, if the relations among facial features are disrupted by inversion, the individual features are perceived as separate parts and are not integrated into a unified whole. As a result, inverted faces elicit neural activity with shorter integrative steps, which we report in the present study.

According to our model, maturation was associated with an increasing number of assimilative steps for inverted compared to upright faces in right parahippocampal gyrus, an area where activation is commonly observed in face recognition and face imagery paradigms for adults (Chua et al. 2007; Powell et al. 2005; Ishai et al. 2002; Bernard et al. 2004; Hung et al. 2010), as well as face recognition in children (Taylor et al. 2011b). Thus, the prominent interaction between development and face inversion displayed by this region may be due to the fact that participants had to judge whether each stimulus had appeared on the previous trial. This result points to an increasing involvement of parahippocampal structures in face recognition.

Our data also revealed a salient transition from childhood/adolescence to adulthood. In adulthood, when inverted faces were presented, neural activity in left inferior prefrontal cortex was associated with additional integrative steps - an effect that was not observed in any other group. A number of studies have noted that this region is sensitive to situations in which face-name associations are formed (Herholz et al. 2001; Sperling et al. 2003; Tsukiura and Cabeza 2008) as well as face recognition (Bernard et al. 2004). This suggests that development of face processing skills may culminate in the emergence of left inferior prefrontal cortex as a unique conduit in the fully mature brain that facilitates the recognition of inverted faces via additional integrative stages.

The fact that inverted faces elicited additional integrative stages further supports the notion that they are perceived “by parts”. If each feature of an inverted face is perceived separately, then more discrete stages of integration should be required compared to upright faces, which are consolidated and perceived holistically. To our knowledge, ours is the first study to operationalize information integration and to
demonstrate differences in how upright and inverted faces are perceived from this perspective.

Why were face inversion effects for the rate of integration localized to inferior temporal cortices, while the effects for the number of integrative steps were localized to “higher” areas, such as the parahippocampal gyrus and prefrontal cortex? We speculate that this may reflect the hierarchical positions of these regions in the context of object recognition. At the level of inferior temporal cortex, representations of complex visual stimuli such as upright faces and features of inverted faces (e.g. eyes, mouth, etc.) are formulated by assimilating low-level visual features from the receptive fields of neurons in striate cortex. Compared to an inverted face, the representation of an upright face must be endowed with additional information about the configuration of individual features, as well as their overall, holistic arrangement, resulting in longer integrative steps.

The present study used a working memory recognition task that required participants to compare the representation of each presented stimulus to the working memory representation of the stimulus presented in the previous trial. Therefore, the involvement of parahippocampal and prefrontal cortex may reflect this secondary cognitive requirement. For upright faces, the comparison is presumably made between two holistic representations. Conversely, inverted faces necessitate a piecemeal comparison between multiple parts of each face, resulting in a greater number of integrations.

3.4.4 Methodological considerations

The strength of our analysis depends on two key questions. First, is the present method of delineating output events physiologically valid? A long-standing assumption is that the basic units of information transfer in the central nervous system are action potentials transmitted by single neurons, while our method is based on electromagnetic recordings that reflect the activity of entire populations of neurons. Therefore, peaks in MEG signal power obviously do not carry information in the same way as action potentials. Nevertheless, neuron soma constitute the grey matter nodes of brain networks and in that sense, post-synaptic potentials represent outputs from these nodes. Peaks in signal power are relevant output events because they originate from synchronized post-synaptic potentials.
Second, is our interpretation of gamma-distributed inter-event times the most plausible? There exist a number of attractive conceptual alternatives based on diffusion models. For example, in cognitive psychology reaction times can be modeled as biased random walkers (biased by incoming information) crossing a “threshold” boundary (Ratcliff 1979; Van Zandt and Ratcliff 1995). Diffusion is also commonly used to model spiking in single neurons (Smith and Ratcliff 2004; Codling et al. 2008). In integrate-and-fire models, inter-spike intervals can be thought of as the time required for a random walker to reach an absorbing boundary. In the context of this experiment, it may be possible to formulate such a diffusion model, parameterized by starting position and bias (i.e. rate of information flow). Given the nature of our data, both integration and diffusion seem to be plausible generative mechanisms and the way in which diffusion models the accumulation of information is similar in spirit to the integrative process we have focused on.

### 3.4.5 Conclusion

By isolating local maxima in the scalogram and treating the time between successive events as a stochastic output process, we have been able to conceptualize neural activity in terms of assimilation of information. We have shown that changes in how information is integrated and emitted manifests as a nonstationarity of inter-event times, that can be meaningfully related to specific cognitive processes and their development. Here we have shown that this model can help to relate the development of face processing skills (and how they are disrupted by face inversion) to information processing and integration. Our approach demonstrates that inversion fragments face processing, resulting in a greater number of shorter integrative episodes. Our data shed new light and provide further evidence for the well-known holistic and configural accounts of face processing.
Chapter 4

Communication efficiency in brain networks

adapted from Mišić et al. (2014) PLoS Comput Biol

4.1 Introduction

Constrained by finite resources, such as metabolism and physical space, which place severe limits on the number and density of synaptic connections, brain networks are an example of how optimized topology may facilitate information flow. The structural topology of cortical networks can be represented and formally studied using the graph model (Albert and Barabási 2002; Boccaletti et al. 2006; Easley and Kleinberg 2010), whereby the brain is spatially parcellated into a set of grey matter nodes interconnected by a set of white matter edges (Rubinov and Sporns 2010; Bullmore and Sporns 2009; Stam and van Straaten 2012). This approach has revealed several aspects of network organization that theoretically confer an increased capacity for information processing, including small-world connectivity (Hilgetag et al. 2000a; Sporns and Zwi 2004; Gong et al. 2009), the presence of hubs (Sporns et al. 2007) and cores (Hagmann et al. 2008), cost-efficient spatial embedding (Bassett et al. 2010; Bullmore and Sporns 2011) and the coexistence of local segregation and global integration (Tononi et al. 1994). Recent studies have also uncovered a “rich club” of hub nodes that are more densely interconnected with each other than predicted by chance (Zamora-López et al. 2010; van den Heuvel and Sporns 2011) and that participate in a disproportionately high number of shortest paths in the network (van den Heuvel et al. 2012; Harriger et al. ...)
The rich club is hypothesized to act as a central backbone for signal traffic, allowing for rapid integration and dissemination of signal traffic (van den Heuvel et al. 2012).

While this graph theoretic approach can articulate the diverse properties of static neural connectivity, it does not take into account the dynamics of information flow on that connectivity. If information flow is introduced into the network, how does neural connectivity influence the efficacy and speed of communication? In other words, how does network topology enable and constrain the capacity of brain networks to globally integrate information? For instance, while certain areas may bridge distant communities and potentially function as hubs by virtue of their connectivity, other areas may be ill-suited as conduits for information transfer because of their position in the network. Under conditions of elevated network traffic such regions could become bottlenecks, imposing limits on the relay of information (Tombu et al. 2011).

To determine the effect of topology on inter-regional communication, we implemented a macaque anatomical brain network as a dynamic system in which units of information flow between grey matter nodes along existing anatomical paths (Fig. 4.1). This allowed us to estimate several metrics of information flow in the network, including the proportion of time a given brain region is in use (utilization), the load on a given brain region (node contents), the time it takes for a unit of information to travel from its source region to its target region (transit time) and the probability of losing information (blocking). The goal of the present study was to use these performance metrics to address the following questions about communication in brain networks. First, does the unique topology of brain networks offer any particular advantage in terms of information processing, and how do brain networks compare to other networks with the same number of nodes and edges, but different topologies? Second, which features of brain network organization contribute most to its capacity for efficient communication? Third, which anatomical regions and pathways are most important for communication?
Figure 4.1: **Discrete-event simulation.** Schematic showing the propagation of two signal units in a simple 3-node, 2-pathway network.


4.2 Materials and methods

4.2.1 Anatomical and reference networks

The anatomical connectivity data set used in the present study was derived from the online Collation of Connectivity data on the Macaque brain (CoCoMac) database, comprised of data from 413 tract tracing studies of the macaque (Stephan et al. 2001; Kötter 2004). The database was originally queried by Modha and Singh (2010) and further refined by Harriger et al. (2012). To facilitate comparison with previous reports, only cortical nodes were included. The final directed network was comprised of 242 nodes and 4090 edges and was fully connected, such that each node maintained at least one incoming and one outgoing edge.

Two populations of control networks - one randomized and one latticized - were generated to explore the extent to which the topology of the macaque connectivity matrix influenced the simulation results. Latticized and randomized networks represent diametrically opposite network configurations. The purpose of generating randomized and latticized versions of the macaque network was to construct a spectrum of topologies around the macaque network. Latticized networks, with high local clustering and long characteristic path lengths occupy one end of the spectrum, while randomized networks, with low local clustering and short characteristic path lengths, occupy the other end. This allowed to address the question of what particular advantages or disadvantages the macaque network has compared to other configurations with the same number of nodes and edges and the same degree sequence.

Randomized networks were generated using a Markov switching algorithm that randomly swapped pairs of edges (Maslov and Sneppen 2002). Latticized networks were generated using a modified version of the same algorithm, whereby the edges were swapped only if they moved closer to the main diagonal as a result (Sporns and Zwi 2004). By randomly re-ordering edges and forcing them closer to the diagonal, the topology of the original network is destroyed and replaced by one where neighbouring nodes are more likely to be connected, as in a ring lattice. Both sets of control networks were degree-matched in the sense that the in-degree and out-degree of each node was preserved. Statistical assessment was performed by comparing 100 simulations on the CoCoMac network with 100 simulations on a randomized null network, for 100 null network realizations. Comparisons between node-specific metrics were made
Figure 4.2: **Small-world network.** The small-world network, proposed by (Watts and Strogatz 1998) begins by connecting nodes with their nearest neighbours to produce a highly regular, lattice-like graph with high clustering and long characteristic path length. Edges are then randomly rewired. When the rewiring probability is high (close to 1), the graph is highly random, with low clustering and a short characteristic path length. However, when the rewiring probability is between 0 and 1, there exists a broad class of networks with dense clustering (similar to a lattice) and short characteristic path length (similar to a random network), termed “small-world” networks. Reprinted by permission from Macmillan Publishers Ltd: *Nature* 1998;393:440-2. Copyright 1998.

using Welch’s $t$-test for samples with unequal variances (Welch 1947), and evaluated with respect to degrees of freedom determined using the Satterthwaite approximation (Satterthwaite 1946). To control the false discovery rate, $p$-values were corrected following the procedure outlined by (Benjamini and Hochberg 1995).

A similar procedure was performed for synthetic small-world (Watts and Strogatz 1998) and rich club (Colizza et al. 2006; van den Heuvel and Sporns 2011; Zamora-López et al. 2010) networks and their respective null models. A network containing a rich-club was created from a random network by endowing a sub-set of the nodes (the rich club) with greater connection density than the rest of the network, and an even greater connection density amongst each other. The randomized and latticized controls were then created as described above. For the small world scenario, the
starting point was a ring lattice. A small world network was generated by randomly
permuting 10% of the edges, while a completely randomized network was generated
by further permuting each edge 100 times (see Fig. 4.2).

4.2.2 Rich club detection

Our results have considerable implication for the rich club feature of brain network
topology and so for completeness we briefly rehearse the procedure for detecting and
defining rich clubs. Fuller descriptions of the rich club phenomenon can be found
elsewhere, for brain networks in general (van den Heuvel and Sporns 2011; van den
Heuvel et al. 2012), as well as for this particular network (Harriger et al. 2012).

For a given graph, a rich club is defined as a set of high-degree nodes (a subgraph)
that are more densely connected amongst each other than would be expected on the
basis of degree alone (Colizza et al. 2006). Rich club classification is made with respect
to a range of node degrees. For a given degree \( k \), all nodes with degree \( \leq k \) are stripped
from the network. A rich club coefficient \( \phi(k) \) is calculated as the ratio of remaining
connections to all possible connections. Thus, \( \phi(k) \) can be thought of as the density of
the subgraph. For the same set of nodes, the ratio is also computed with respect to
10,000 degree-matched randomized networks. The normalized rich club coefficient,
\( \phi_{\text{norm}}(k) = \phi(k)/\phi_{\text{random}}(k) \), measures the density of the subgraph relative to the null
model where the global topology has been destroyed. These steps are repeated for a
range of \( k \), from the lowest to the second-highest degree in the macaque network (2 to
121). A \( \phi_{\text{norm}}(k) \) consistently greater than 1 for a range of \( k \) suggests the existence of
rich club organization.

Therefore, across the range of \( k \) it is possible to define unique sets of rich club nodes
 corresponing to different values of \( k \). These nodes can then be positioned in a nested
hierarchy of rich club “levels”, ranging from those containing nodes with the highest
degree to those containing nodes with the lowest degree. In the present study, we
follow the classification made by Harriger et al. (2012), whereby two rich clubs were
singled out. The first, RC1 \( (k = 12) \), was more densely interconnected and comprised
of fewer nodes, with greater minimum degree. The second, RC2 \( (k = 34) \), was less
densely interconnected and contained more nodes, with smaller minimum degree.
RC2 is a subset of RC1, and by examining these two levels of the rich club, it is possible
to identify robust relationships between rich-club organization and information flow
as estimated by our model. A list of rich club areas and their spatial locations are included in Appendix H.

Once nodes have been classified as either rich club or non-rich club, it becomes possible to classify edges as well. Namely, edges that connect non-rich club nodes to non-rich club nodes are classified as “local”, those connecting non-rich club nodes to rich club nodes as “feeder” and those connecting rich club nodes to other rich club nodes as “rich club”.

4.2.3 Discrete-event simulation

Signal units were generated and introduced in the network according to a Poisson process with rate $\lambda$, i.e. with exponentially distributed inter-arrival times. For each signal unit, a source node and destination node were randomly selected. To reach its destination node, the signal propagated to one of the neighbouring nodes, with equal probability for each. The time spent at each node (service time) was exponentially distributed with rate $\mu = 0.02$. If a signal unit arrived at a node that was occupied, a queue was formed (Fig. 4.1). Units entered the node on a last-come-first-served basis, also known as last-in-first-out (LIFO) queueing (Kleinrock 1972; Banks et al. 1984; Takagi 1993a). A maximum buffer size was imposed ($H = 20$), such that a signal units arriving at a full buffer caused the oldest signal unit in the queue to be ejected and removed from the system. Upon reaching the destination node, the unit was removed from the network. The purpose of queueing is simply to ensure that information flow is interactive, while a finite buffer size allowed us to model imperfect signal transmission (Faisal et al. 2008). Buffer capacity is not a critical parameter, in the sense that it cannot induce a phase transition in the system. Changes in buffer capacity will produce quantitative, but not qualitative, changes in system behavior (Appendix B, Fig. B.1).

Fig. 4.1 illustrates how the simulation works. In this example, there are three nodes, each of which has a two-slot buffer, that are interconnected by two pathways. At time 2, a signal (red) completes service at node 1 and moves to node 2. The signal (red) arrives at node 2 and enters the buffer. Due to the last-come-first-served queueing discipline, the signal moves to the front of the buffer. Meanwhile, the signal (teal) in the buffer at node 1 commences service (time 3). At time 4, the signal (teal) at node 1 completes service and moves to node 2, where it moves to the front of the buffer and
displaces the oldest signal (green), which is ejected from the network (time 5). At time 6, the signal (blue) at node 2 completes service and moves to node 3, where it enters the empty buffer (time 7). Meanwhile, the signal (teal) at the front of the queue at node 2 commences service.

This type of model has two characteristic modes of operation. At low intensities (external arrival rates), the total number of signal units in the network fluctuates around some finite value and the system is said to be in a steady-state. As the intensity is increased, there is a qualitative change (phase transition) in the system dynamics, characterized by a monotonic increase in the number of signal units in the network until all buffers are filled, leading to “jamming” (Boccaletti et al. 2006; Tadić et al. 2004). The key variable is the ratio between the arrival rate and service rate at each node. Therefore, we fixed the service rate ($\mu$) and varied the rate of external arrivals ($\lambda$). The focus of the present study was on the steady-state behavior of the network, and the range of external arrival rates ($\lambda = 0.005, 0.01, 0.015, 0.02$) was chosen to sustain stationary flow, prior to the phase transition.

All simulations were run for 2 million dimensionless time units. Due to the presence of stochastic time variables in the simulation (inter-arrival times and service times), the state of the system was updated at non-uniform time points. Upon completion, the time series of system states were linearly interpolated to produce uniformly sampled time series (Appendix D, Fig. D.1). An initial transient of 40,000 time units, during which the system state had not yet stabilized (determined via the ensemble average method recommended by Banks et al. (1984)), was discarded from further analysis to avoid transitory effects. The Mersenne Twister (Matsumoto and Nishimura 1998) was used to generate a uniform distribution, which was then used to generate exponentially distributed random numbers (inter-arrival times and service times) using the standard inverse transform method. All simulations were implemented in Matlab (Mathworks Inc., Natick, MA) and independently verified in Artifex (RSoft Design Group Inc., Ossining, NY), as well as analytically (Appendix A, Figs. A.1,2).

All signal units were uniquely identified, allowing for their position and complete trajectory in the network to be traced across the simulation. These trajectories were then analyzed to compile a set of node-, edge- and network-level statistics. For each node, we calculated the mean proportion of time the node was busy (utilization), the probability of signal loss (blocking) and the mean system contents. For each edge, we calculated the mean throughput of signal units. For each network, we calculated the
mean utilization and blocking rates across nodes, as well as the total number of signal units successfully transmitted from source to destination (throughput) and the mean latency of those transmissions (transit time).

More formally, simulation variables were defined as follows. A node \( i \) at time \( t \) has two components: the server contents \( s_i(t) = \{0, 1\} \), which describes the number of signal units currently in service, and the queue length \( q_i(t) = \{0, ..., H\} \), which describes the number of signal units waiting in the buffer. The node contents \( n_i(t) \) were thus defined as

\[
n_i(t) = s_i(t) + q_i(t).
\] (4.1)

Likewise, the contents at any existing channel from node \( i \) to node \( j \) was \( c_{ij}(t) \). The total network load \( N(t) \) is then the sum of all node and channel contents:

\[
N(t) = \sum_i n_i(t) + \sum_i \sum_j c_{ij}(t).
\] (4.2)

The utilization of node \( i \) is the proportion of simulation time during which \( s_i = 1 \). The blocking probability at node \( i \) was calculated as the number of signal units ejected from \( i \) divided by the total number of signal units arriving at \( i \).

The total time a signal unit spends at a single node, \( T \), is the sum of the waiting time in the queue \( T_w \) and the service time in the node \( T_s \)

\[
T = T_w + T_s.
\] (4.3)

Both \( T_w \) and \( T_s \) are stochastic processes, with \( T_w \) determined by the the topology and dynamics on the network, while \( T_s \) is drawn from an exponential distribution with rate \( \mu = 0.02 \). For any signal unit, the transit time is the sum of waiting and service times across all nodes traversed from source to destination. Transit time statistics are calculated only for signals that successfully reached their destination.

4.3 Results

4.3.1 Network statistics

Measures of information flow on the macaque network were compared against a spectrum of degree-matched control networks with an equal number of nodes and
edges, but systematically altered topology. One set was comprised of randomized networks, while the other set was comprised of latticized networks. Under conditions of increasing load (see Appendix G, Fig. G.1), all networks experienced increased blocking and utilization, as well as decreased throughput, thus exhibiting signs of congestion. Mean transit times for signal units that reached their destination also decreased with increasing load, but this counterintuitive observation is the result of decreased throughput. At lower network load, more signals reach their destination but some may take a long time to do so, which increases the mean transit time. As the network becomes congested, many such signals may get dropped at over-utilized nodes and never reach their destination, and thus cannot influence the mean transit time.

The macaque network was intermediate on all information flow statistics compared to the randomized and latticized networks. This is consistent with the fact that the randomized and latticized networks represent two extreme and diametrically opposite network configurations and suggests that the organization of the macaque network serves to strike a balance between speed, reliability, utilization and total throughput. Compared to its randomized control network, information flow on the macaque network was characterized by significantly higher loss rates, faster transit times and lower throughput (Fig. 4.3 top, \( p < 0.05 \) for all measures, Appendix F, Tables F.1,2,3), suggesting that neural connectivity may be optimized for speed rather than fidelity. In general however, the two networks performed similarly, while the latticized control network performed much differently, with significantly lower utilization and throughput, shorter transit times and near total loss of information (\( p < 0.05 \) for all measures).

I next sought to determine which feature of network topology most contributed to this pattern of results. To investigate the degree to which the existence of a rich club influences information flow, a synthetic network containing a rich club was created, as well as a set of degree-matched randomized and latticized controls (Appendix E, Fig. E.1). Information flow on these networks was contrasted with a canonical small world network, which is a ubiquitous and well-studied model for many different kinds of information-processing networks, including neural networks (Watts and Strogatz 1998). The pattern of results produced by the synthetic small world and rich club networks and their respective randomized and latticized controls were considerably different (Fig. 4.3, middle and bottom). Importantly, the system statistics associated with the rich club network were nearly identical to the macaque network (see Ap-
Figure 4.3: **System-level statistics.** Simulations were run for three different scenarios: CoCoMac brain (top), small-world network (middle) and rich-club network (bottom) and their respective randomized (green) and latticized (red) control networks. For each network, the mean transit time, utilization, blocking probability and throughput are plotted at four different intensities. For the starting networks in each scenario (CoCoMac, lattice and rich-club), the curves represent the average over 100 simulations. For the randomized/latticized versions of each network, the curves represent the average across 100 simulations on 100 realizations of each network.

The results observed for the macaque network were also similar to the canonical Watts-Strogatz small world network (Watts and Strogatz 1998), but to a significantly lesser extent. Overall, this suggests that the rich club is an important topological feature for information flow in the brain, as defined by these four statistics.

### 4.3.2 Node statistics

We now examine regional contributions of the CoCoMac network in detail. To study the individual relevance of nodes for information flow, three complementary node-
level metrics of congestion were used: utilization, blocking and mean node contents. Information flow was highly heterogeneous across the network, with some nodes vulnerable to overwhelming influx, while others experienced only occasional traffic (Fig. 4.4). To a large extent congestion at a given node was predicted by the number of afferent projections to that node (in-degree, $r^2 = 0.83, 0.15$ and $0.45$ and $p = 1.2 \times 10^{-93}, 2.0 \times 10^{-11}$ and $1.8 \times 10^{-33}$ for utilization, blocking and node contents, respectively) and this is expected given the fact that in the present model information flow is implemented as an interactive random walk (Boccaletti et al. 2006; Tadić et al. 2004; Duch and Arenas 2006). With the exception of CA1, all nodes with the largest average contents were previously identified as part of the rich club (Harriger et al. 2012)(Appendix H, Fig. H.1), indicating that membership in this densely inter-connected subgraph entails a heavy workload. Much of the congestion appears to be concentrated at three distinct sites, mainly along the medial surface, roughly corresponding to medial prefrontal cortex, medial/inferior temporal cortex and precuneus/posterior cingulate cortex.

To determine the extent to which these congestion metrics depend on topology rather than degree sequence, we statistically compared them to a set of metrics from simulations run on a population of randomized control networks for which the topology had been altered while preserving degree sequence (Maslov and Sneppen 2002). Fig. 4.5A shows the “raw” mean differences between the two networks for the contents of each node, while Fig. 4.5B shows the spatial distribution of these differences. Due to the high level of consistency between the three metrics of congestion (utilization, blocking and contents), only the results for node contents are shown. Nodes with contents that are significantly different ($p < 0.05$, controlled for multiple comparisons using false-discovery rate correction) for the two sets of the networks are labeled. Interestingly, while the majority of these nodes are part of the rich club, nearly half experience greater congestion in the macaque network, while half experience greater congestion in the randomized networks. This suggests that macaque cortical connectivity imposes a characteristic set of traffic patterns, such that signal traffic is directed towards some nodes and away from others, in contrast to what would be expected based only on the degree of these nodes.
Figure 4.4: **Node-level statistics.** Top: Utilization, blocking probability, node contents and in-degree are shown for each of the 242 nodes on the CoCoMac network, averaged over 500 simulations ($\lambda = 0.01$, $H = 20$, $\mu = 0.02$). For illustrative purposes, nine nodes with the highest node contents are labeled. Bottom: Inflated surface renderings showing the anatomical distribution of each statistic, for the medial and lateral surfaces.

### 4.3.3 Edge statistics

We next consider information flow with respect to specific edges. Given the relevance of the rich club in the macaque network (Harriger et al. 2012), we classified edges according to whether they connect rich club nodes (van den Heuvel et al. 2012). Edges connecting two non-rich club nodes were classified as (L)ocal, those connecting a non-rich club and a rich club node as (F)eeder and those connecting two rich club nodes as (R)ich club. Moreover, these classifications were made with respect to two rich club levels, RC1 and RC2, which represent a more conservative and a more liberal definition of the rich club (Harriger et al. 2012). An initial observation is that this stratification of edges closely resembles the patterns of edge throughputs. In particular, projections with greater throughput appear more likely to be those connected to at least one rich
Figure 4.5: **Node contents for the CoCoMac and randomized control networks.**

(A) The mean differences in node contents between 100 simulations based on the CoCoMac network, and 100 simulations on 100 randomized networks ($\lambda = 0.01, H = 20, \mu = 0.02$). Nodes with statistically significant differences are labeled. (B) Surface renderings of the mean differences in node contents.

club node, i.e. Rich Club or Feeder. Despite the fact that the vast majority of edges in the macaque network are Local, followed by Feeder and then Rich Club (Harriger et al. 2012), the mean throughput per edge is greatest for Rich Club edges, followed by Feeder and Local (Fig. 4.6B). In other words, traffic tends to concentrate not just
at rich club nodes, but also at the edges around them, effectively encompassing their local neighborhoods.

Figure 4.6: **Edge throughput.** The mean number of signal units carried by each edge. Edges are classified as Rich Club (R), Feeder (F) or Local (L). Results are shown for 500 simulations ($\lambda = 0.01$, $H = 20$, $\mu = 0.02$) on the CoCoMac network.

### 4.3.4 Path statistics

Finally, we investigate information flow with respect to every possible pair of source and target nodes. For each pair, all completed trajectories are compiled in order to compute the total number of deliveries (throughput), as well as their mean transit time or delay. For both the throughput and transit time statistics, taking the mean across sources results in greater variance than taking the mean across targets (Fig. 4.7A,B). For the target nodes, both statistics showed substantial association with in-degree ($r^2 = 0.25$ and $0.86$, $p = 6.1 \times 10^{-21}$ and $7.8 \times 10^{-106}$ for transit time and throughput, respectively). In other words, the mean throughput and transit time were much more dependent on the destination, rather than the source, indicating that some nodes in the network are intrinsically easy to reach, while others are intrinsically difficult.

A non-monotonic relationship emerges when comparing the mean throughput and the mean transit time across target nodes (Fig. 4.7C, dark grey). When the total throughput is low, any increase in throughput results in slower transit times. However, for a subset of nodes with a high throughput this relationship does not hold and these nodes tend to receive information much faster than would be expected. Most of these nodes belong to the rich club (RC2), indicating that rich club nodes receive more information than other nodes in the network, and do so with a disproportionately
Figure 4.7: **Delays and deliveries.** Results are shown for 500 simulations on the CoCoMac network (A,B,C) and 1 simulation on 500 randomized networks (C) ($\lambda = 0.01, H = 20, \mu = 0.02$). Means of transit times (A) and throughput (B) are either taken across all target nodes (showing the mean for individual source nodes), or across source nodes (showing the mean for individual target nodes). (C) The relationship between transit time and throughput across all target nodes, with rich club (RC) nodes highlighted.
faster latency. A similar relationship is observed for degree-preserving randomized controls (Fig. 4.7C, light grey), indicating that the effect is largely due to the high degrees of rich club nodes. Rich club connectivity enhanced the effect. As expected from Fig. 4.3, the randomized controls generally have slightly higher throughput, but also longer transit times. This was particularly true for rich club nodes, which received significantly fewer signal units when embedded in the macaque than in randomized networks ($p < 0.05$ for RC1 and RC2), but did so with significantly faster transit times ($p < 0.05$ for RC1 and RC2).

### 4.4 Discussion

The complex anatomical connectivity of the central nervous system suggests that inter-regional communication is important for the functioning of the brain, and in the present report we systematically investigated the effect of network topology on communication. Utilizing a modeling paradigm from telecommunications and statistical physics (Boccaletti et al. 2006; Duch and Arenas 2006), we superimposed a dynamic system on an empirically derived network describing macaque cerebral cortex. Our results highlight multiple ways in which structural connectivity has the potential to exert considerable influence on information flow in brain networks.

In terms of global information flow statistics, the macaque network was found to be intermediate to its latticized and randomized reference networks, mirroring the notion that the complex topology of structural networks represents a trade-off between wiring cost and communication efficiency for integrative processing (Bassett et al. 2010; Bullmore and Sporns 2011; Goñi et al. 2013). In particular, the macaque network exhibited an economic balance between speed, fidelity, utilization and sheer volume of transmission. Compared to degree-matched random networks, the macaque network appeared to prioritize speed of transmission over throughput and reliability.

Although many studies have reported evidence of small world organization in structural (Hilgetag et al. 2000a; Sporns and Zwi 2004; Gong et al. 2009) and functional networks (Bassett et al. 2006; Rubinov et al. 2009a), a canonical small world model by itself could not account for the information processing characteristics observed in the macaque network. However, the added presence of a rich club largely replicated the information flow signature of the macaque network. Therefore, our data suggest
that the small world property (in the Watts-Strogatz sense (Watts and Strogatz 1998)), together with the rich club, is necessary to produce the macaque-like pattern, but by itself is not sufficient. Several recent studies have postulated that a densely interconnected rich club has the potential to facilitate global integration by providing an easily accessible high-capacity backbone that serves to attract and disseminate interregional signal traffic (Gómez-Gardeñes et al. 2010; Zamora-López et al. 2010; van den Heuvel and Sporns 2011; van den Heuvel et al. 2012; Harriger et al. 2012). By demonstrating that the rich club is a principal topological feature with respect to communication dynamics, our model lends further support to this notion.

Forming a triangle spanning frontal cortex, posterior cingulate cortex/precuneus and medial temporal cortex, the rich club subgraph proved to be a prominent axis in the information processing architecture of the network. Rich club nodes, as well as connections involving rich club nodes, absorbed the greatest signal traffic, indicating that the rich club of densely connected hubs supports the capacity to efficiently centralize and, presumably, integrate information. The fact that rich club nodes are more likely to exhibit signs of congestion warrants further investigation into their potential role as bottlenecks in information processing. Several empirical studies have reported evidence of bottlenecks limiting processing capacity in attention (Tombu et al. 2011) and response selection (Jiang and Kanwisher 2003), and the areas they implicate show considerable correspondence with rich club regions (including medial prefrontal cortex and precuneus), although a direct comparison between the macaque network and human fMRI studies is difficult.

Despite the fact that rich club nodes were among the most congested, comparisons with randomized networks revealed that rich club connectivity may serve to shape information flow, whereby signal traffic is biased towards some nodes and away from others. While a number of rich club nodes consistently experienced heavier traffic than would be expected on the basis of their degrees, others consistently experienced lighter traffic than would be expected on the basis of theirs. Why macaque network topology, and the rich club in particular, shapes cortico-cortical communication in a way that imposes this specific pattern of information flow, remains unclear. It is noteworthy that the under-congested nodes are areas associated with making eye movements, tracking and acting towards objects in space and fusing visual and proprioceptive information. Many of these areas are part of the dorsal attention sub-network (Hopfinger et al. 2000; Kastner et al. 1999; Corbetta et al. 2002), which presumably must be continually
responsive and capable of rapidly integrating and communicating information. We therefore speculate that the topology of the global network is configured in a way that relieves congestion at these dorsal attention areas to facilitate fast and efficient interaction with the external environment.

Previous analyses have shown that, despite constituting only a small part of total network density, rich club connections participate in the greatest number of shortest paths in the network (van den Heuvel et al. 2012; Harriger et al. 2012). This has led to the hypothesis that if rich club topology is configured in a way that facilitates cortico-cortical communication via shortest paths, there may exist a set of routing or navigation strategies to take advantage of this feature (van den Heuvel et al. 2012), such as “greedy” routing (Boguna et al. 2008). However, our data demonstrate that the rich club is central to global communication even if information flow is governed by simple diffusion rather than shortest path communication, potentially eschewing the need for a more complex routing mechanism (da Fontoura Costa and Sporns 2006; da Fontoura Costa et al. 2007; Dimitriadis et al. 2012). To characterize the organizational principles of global information flow, further investigation is necessary to determine which types of routing strategies can best take advantage of the unique connectivity of brain networks, as well as which types of routing strategies best replicate empirical functional data.

Regarding the modeling paradigm employed in this study, we underscore that our primary goal is to investigate how pairs of nodes communicate in the context of the whole network, with discrete signal units representing the ability of brain regions to influence one another. This simplifying assumption allows us to trace the trajectory of each signal unit as it propagates in the network, and hence to calculate various metrics about the potential for communication that is afforded by the anatomical connectivity. External arrivals represent the assumption that new information is continuously generated and communicated in the network. The source of this information may be either stimulation exogenous to the nervous system, or some endogenous process. Poisson arrivals were chosen because at the level of individual neurons, inter-spike intervals are found to be exponentially distributed (Tuckwell 2005) and likewise, in psychophysics and signal detection, the Poisson process is often used to model stimulus fluctuations and other statistical properties of the sensory environment (Barlow 1956). Queues and finite buffers allow us to model how network topology constrains information flow. Queueing is a mechanism by which signal units are made to in-
interact as they flow through the network, modeling the interplay between multiple information flows on top of the structural network (Liu 1996). Finite buffers allow for the possibility of signal loss, modeling the poor fidelity of neural transmission (Faisal et al. 2008).

The strength of the modeling approach pursued here is that it allows one to generate relative metrics about network communication. The approach is complementary to other, more physiologically realistic paradigms for modeling global system dynamics (Honey et al. 2007, 2009; Ghosh et al. 2008b; Deco et al. 2009, 2011; Jirsa et al. 2010), which do not model information transmission directly. These models suggest that three key ingredients are needed to generate realistic brain dynamics: empirically derived patterns of structural connectivity, time-delayed transmission and noise (Deco et al. 2011; Jirsa et al. 2010). Indeed, a queueing network model has the potential to incorporate all three, and the present implementation includes both empirically derived connectivity and stochastic dynamics.

Altogether, our results reveal a dynamic aspect of the global information processing architecture and the critical role played by the so-called “rich club” of hub nodes. Our work lays the foundation for further systematic study of organizational principles for communication in large-scale brain networks, including routing strategies and resource allocation.
Chapter 5

A convergence zone

adapted from Mišić et al. (in preparation)

5.1 Introduction

The hippocampal formations are among the most studied and best understood areas of the brain. Along with adjacent cortical structures, such as the entorhinal, perirhinal and parahippocampal cortices, the hippocampi are thought to facilitate memory function, particularly the initial encoding of memories (Squire 1992; Tulving and Craik 2000; Corkin 2002; Moscovitch et al. 2005). Lesion studies in various model organisms and in man, as well as functional neuroimaging studies, have consistently found that the hippocampal formations are specialized for forming conjunctions between arbitrarily different elements.

The functional anatomy of the brain supports the notion that the hippocampi may be central structures that serve to bind together distributed sites in neocortex to represent a memory. Sensory information converges upon the hippocampi via multisynaptic projections, such that all fields of the hippocampi receive polysensory input from association areas of the neocortex, via perirhinal and parahippocampal cortices. In particular, the final field of the hippocampi, CA1, receives multimodal input. A prominent idea is that the hippocampi are a “convergence zone” in the brain, whereby successive levels of convergence culminate in maximally integrative regions, such as medial temporal lobe cortices (Damasio 1989; Mesulam 1998). Altogether, the literature suggests that the hippocampi should occupy an important position in the whole-brain network.
However, despite the prominent role of the hippocampi in memory function, quantitative analyses of anatomical and functional whole-brain networks have largely failed to demonstrate the topological centrality of the hippocampi. Recent graph-theoretic studies of the connectome, both in man (Hagmann et al. 2008; van den Heuvel et al. 2012; Gong et al. 2009) and in the macaque (Sporns et al. 2007; Honey et al. 2007; Modha and Singh 2010; Harriger et al. 2012) have consistently found that the hippocampi are unlikely to be a hub in the context of the whole brain, as they are not among the most highly connected areas, nor do they appear to occupy a position along many of the shortest paths in the network. Likewise, the hippocampi are not highly central in large-scale functional networks either (Buckner et al. 2009; Tomasi and Volkow 2010; Honey et al. 2009). Even large-scale computational models have found no evidence to suggest that the hippocampi are central in the context of the whole-brain connectome (Honey et al. 2007; Ghosh et al. 2008b).

However, these studies of the brain have largely focused on the connectivity of the network, rather than on how that connectivity supports communication. Results presented in the previous chapter (Chapter 4) suggested that the final field of the hippocampi, CA1, may be important for communication, as it was the only area outside of the rich club that was significantly over-congested. In the present study I show that when the communication capacity of the whole-brain network is taken into account, the hippocampus becomes a critical area, above and beyond its ostensibly average topological attributes.

5.2 Materials and methods

Data were generated according to the procedure described in Chapter 4. For completeness, the essential details of the procedure are included here. Please note that, as in chapter 4, the structural connectivity matrix represents a single hemisphere of the macaque brain.

5.2.1 Anatomical and reference networks

The anatomical connectivity data set used in the present study was derived from the online Collation of Connectivity data on the Macaque brain (CoCoMac) database, comprised of data from 413 tract tracing studies of the macaque (Stephan et al. 2001; Kötter
2004). The final directed network was comprised of 242 nodes and 4090 edges and was fully connected. Importantly, the connectivity matrix contained several nodes that are part of the hippocampal formation and neighbouring cortical structures, including parahippocampal areas TF/TH, perirhinal areas 35/36, entorhinal cortex, dentate gyrus, subfield CA3, subfield CA1 and subiculum.

A population of randomized surrogate networks was generated to explore the extent to which the topology of the macaque connectivity matrix influenced the simulation results. Randomized networks were generated using a Markov switching algorithm that randomly swapped pairs of edges (Maslov and Sneppen 2002). Statistical assessment was performed by comparing 100 simulations on the CoCoMac network with 100 simulations on a randomized surrogate network (serving as a “null” network), for 100 network realizations.

5.2.2 Discrete-event simulation

Signal units were generated and introduced in the network according to a Poisson process with rate $\lambda$, i.e. with exponentially distributed inter-arrival times. For each signal unit, a source node and destination node were randomly selected. To reach its destination node, the signal propagated to one of the neighbouring nodes, with equal probability for each. The time spent at each node (service time) was exponentially distributed with rate $\mu = 0.02$. If a signal unit arrived at a node that was occupied, a queue was formed. Units entered the node on a last-come-first-served basis, also known as last-in-first-out (LIFO) queueing (Kleinrock 1972; Banks et al. 1984; Takagi 1993a). A maximum buffer size was imposed ($H = 20$), such that a signal units arriving at a full buffer caused the oldest signal unit in the queue to be ejected and removed from the system. Upon reaching the destination node, the unit was removed from the network.

The service rate ($\mu$) and the rate of external arrivals ($\lambda$) were fixed at $\mu = 0.02$ and $\lambda = 0.01$. The parameters were chosen to according to the findings presented in Chapter 4, i.e. to sustain stationary flow. The results of the present analysis hold for a wide range of external arrival rates ($\lambda = 0.005, 0.01, 0.015, 0.02$), but for simplicity only one data point is shown.

All simulations were run for 2 million dimensionless time units. Due to the presence of stochastic time variables in the simulation (inter-arrival times and service times),
the state of the system was updated at non-uniform time points. Upon completion, the
time series of system states were linearly interpolated to produce uniformly sampled
time series (SI Text Section 6, Fig. S6). An initial transient of 40,000 time units, during
which the system state had not yet stabilized (determined via the ensemble average
method recommended by Banks et al. (1984)), was discarded from further analysis to
avoid transitory effects. The Mersenne Twister (Matsumoto and Nishimura 1998) was
used to generate a uniform distribution, which was then used to generate exponentially
distributed random numbers (inter-arrival times and service times) using the standard
inverse transform method.

5.3 Results

The first step of the analysis briefly recapitulates some of the results presented in
the previous chapter, but with a focus on CA1. Notably, CA1 experiences a high
throughput of signal traffic (Fig. 5.1). Fig. 5.1A shows the complete information flow
profile for the network, with mean utilization, node contents and total arrivals at each
node, while Fig. 5.1B shows the spatial distribution of total arrivals. For all three
metrics, CA1 is ranked #7 out of 242 nodes, placing it in the top 3%.

To determine whether the high ranking of CA1 is due to its degree or due to its
embedding in the global macaque topology, the data from the macaque network were
compared against a “null” model, comprised of randomized surrogate networks. In
these surrogate networks the degree sequence is preserved but the global topology is
destroyed by randomization. Critically, when the network is randomized, informa-
tion flow through CA1 is greatly reduced (Fig. 5.2). The extent to which the role of
CA1 differs when embedded in a randomized network versus the macaque network
can be quantified and statistically assessed by expressing the mean of the macaque
network distribution (red) as a z-score relative to the randomized null distribution.
In the present case, the scores for the node contents, arrivals and utilization were
\( z = 16.63, 8.60 \) and \( 8.52 \) respectively, corresponding to \( p << 0.001 \).

Thus, comparison to a null model confirms that the convergence of signal traffic
at CA1 is not due to its degree, but some other higher-level feature of macaque con-
nectivity. Figs. 5.3A and B confirm this, showing that CA1 behaves in a unique way.
While greater in-degree (i.e. number of afferent projections) is associated with greater
Figure 5.1: **Node metrics.** (A) Three local metrics of communication efficiency (utilization, node contents and arrivals) and information flow are shown for all 242 nodes of the network, averaged over 500 simulations ($\lambda = 0.01, H = 20, \mu = 0.02$). (B) Inflated surface renderings showing the anatomical distribution for the node contents statistic, for the medial and lateral surfaces.
Figure 5.2: **Role of network topology.** Histograms showing the distribution of CA1 node metrics (utilization, node contents and arrivals) for multiple simulation runs. The histograms represent 100 simulations on the macaque network (red) and a single simulation for 100 randomized surrogate networks (blue).
signal traffic, CA1 is a clear outlier. Namely, CA1 attracts signal traffic to an extent that is above and beyond what would be expected on the basis of its in-degree alone.

It is possible that the communication profile of CA1 is not determined by its own degree properties, but by the connectivity in its local neighbourhood. Given the diffusive dynamics of the present model, it may be that signal traffic converges to CA1 because of the nodes that project to it (its in-neighbours). One possibility is that the in-neighbours of CA1 collectively have a higher than average in-degree, and that CA1 is statistically more likely to experience higher levels of signal traffic. A second possibility is that the in-neighbours of CA1 collectively have a lower than average out-degree, thus funneling signal traffic to CA1. The plots in Figs. 5.3 C-F explore these possibilities and suggest that neither is likely, because neither the mean in-degree nor out-degree of the in-neighbours of CA1 is able to explain the high levels of signal traffic at that node. Furthermore, the assortativity plots in Fig. 5.4A and B suggest that the in-neighbours of CA1 have neither higher than expected in-degree nor lower than expected out-degree.

However, it may be possible that the mean connectivity profile of the in-neighbours of CA1 obscures the contribution of individual projections. For instance, it may be the case that there is a small number of projections terminating in CA1 that carry high levels of signal traffic. The total throughput along each directed connection in the macaque network is shown in Fig. 5.5A, while Fig. 5.5B shows the top 10 most traversed connections (i.e. highest-valued elements in the matrix). Note that the two most traversed connections in the network are from TFM to CA1 and from TFL to CA1. TFM and TFL represent the medial and lateral portions of parahippocampal area TF, and are known to project directly to CA1 (Suzuki and Amaral 1990). The histogram in Fig. 5.5C helps to get a sense of the contribution of these two projections. The two projections are not only the most traversed - they are far removed from the distribution, representing extreme outliers.

The most plausible reason why these two projections carry so much signal traffic is because of a severe degree imbalance. Namely, both TFM and TFL have relatively large in-degrees (34 and 41, respectively), and relatively low out-degrees (1 and 2, respectively) (Fig. 5.5D). Thus, TFM and TFL absorb high levels of signal traffic but - as they project only to CA1 and one other node - create a funneling effect, resulting in a convergence of signal units at CA1. This degree imbalance is illustrated in Fig. 5.6, which shows the connectivity between TFM/TFL and CA1. Both TFM and TFL have
Figure 5.3: **CA1 as a communication outlier.** Communication metrics (node contents and arrivals) are compared to connectivity metrics, including in-degree (A,B), neighbours’ mean in-degree (C,D) and neighbours’ mean out-degree (E,F). In panels C-F, “neighbours” refers to nodes that project to CA1.
very large in-degrees but very low out-degrees, causing traffic to be funneled towards CA1.

To investigate whether the directionality of cortical projections contributes to the high signal traffic at CA1, the directions of all projections in the network were reversed by transposing the original CoCoMac adjacency matrix (Fig. 5.7). Thus, all projections \(i \rightarrow j\) were altered to \(i \leftarrow j\). In the new network, the congestion at CA1 disappears, indicating that the directionality of all projections in the network serves to funnel signal traffic to CA1.

### 5.4 Discussion

The present results demonstrate that to assess the functional capacity of the brain, it is important to consider communication dynamics, above and beyond static connectivity. Importantly, the CA1 field of the hippocampus appears to be a critical node, embedded in the connectome in a way that allows signal traffic to converge from multiple distributed areas.

These results are consistent with the notion that the hippocampal formations are a convergence zone for multiple information streams, giving rise to polysensory, multimodal representations (Man et al. 2013; Meyer and Damasio 2009). Our data suggest an integrative role, wherein information from diverse afferents is pooled and presum-
Figure 5.5: **Degree imbalances.** (A) The total number of signal units that traversed a particular connection. (B) The ten most traversed connections. (C) A histogram of all connections in (A), showing the distribution of signal traffic on all connections. (D) The relationship between in-degree and out-degree for all nodes in the network.
Figure 5.6: **Neighbourhood of CA1.** Nodes TFM and TFL have large in-degrees and low out-degrees, causing traffic to be funneled towards CA1. The nodes are spatially positioned in a way that coincides with the directionality of edges, i.e. information is projected from top to bottom.
Figure 5.7: **Role of directionality.** Histograms showing the distribution of CA1 node metrics (utilization, node contents and arrivals) for multiple simulation runs. The histograms represent 100 simulations on the macaque network with the original directions (red) and 100 simulations on the macaque network with reversed directions (blue).
ably assimilated to engender a coherent, multimodal representation. Although the functional anatomy is highly suggestive of this role, to our knowledge these data are the first quantitative topological evidence of such an organizational principle.

The information processing capacity of the hippocampi has traditionally been studied at the local level, with a focus on information flow and plasticity in the hippocampal formations and their local neighbourhood (Battaglia et al. 2011). Our results suggest the hippocampi are also critical for information processing at the global level and that they are central hubs in the context of large-scale networks. Recent research suggests that information processing in hippocampal circuits is mediated by endogenous theta rhythms, with distinct phases in every field of the hippocampi (Goutagny et al. 2009; Dupret et al. 2013). Information flow is then dynamically “routed” by a gamma rhythm riding on the theta troughs, centered at CA1 (Colgin et al. 2009). The results of the present study also point to CA1 as the critical field of the hippocampal formations, but more research is necessary to relate local oscillatory activity to global information flow.

Our data also indicate that the connectivity between hippocampi proper and adjacent cortex in the ventral temporal lobes is particularly important for information flow. In particular, the parahippocampal and perirhinal cortices are thought to be the primary route by which information is exchanged between the hippocampal formations and the neocortex. In the present study, the highest volume of signal traffic arrived to CA1 via direct projections from parahippocampal area TF rather than along the perforant path. This suggests that parahippocampal cortex (TF/TH) may act as gateway for information flow to the hippocampi. The anatomical connectivity of parahippocampal cortex also supports this notion, with many neocortical afferents converging on this particular site (Sewards 2011).

5.4.1 Methodological limitations

In the present study communication dynamics are modeled in terms of diffusive signal traffic. This approach confers several advantages, including the ability to trace individual signal units, as well as the no strong assumptions about the routing and transformation of information in the network. However, the model is also limited in some ways and it is important to consider to what extent the assumptions of the model limit the conclusions drawn from the data.
First, all areas and projections are assumed to have the same capacity, despite the fact that the former vary in size and the latter vary in fiber density. The model was configured in this way to avoid making even stronger assumptions about how size influences processing capacity. For example, it is unclear a priori whether the information processing capacity of an area should vary in a linear or nonlinear fashion with size. In addition, the current configuration of the model makes the results of the study comparable to conventional graph-theoretic analyses, which also assume no variation across nodes or across edges (Stam and van Straaten 2012; Bullmore and Sporns 2009; Rubinov and Sporns 2010).

Second, all signal units are assumed to carry the same amount of information and they flow through the network unchanged. This is unlikely to be the case in real brain networks, where information is fundamentally transformed at each node. The model was configured in this particular way to reduce the complexity of the problem and to focus on the main experimental question: how does the topology of the network influence information flow to the hippocampus? To trace the trajectory of individual signal units it is necessary that they remain unchanged. Thus, discrete signal units merely represent the ability of brain areas to influence each other via anatomical projections.

5.4.2 Conclusion

The present study demonstrates that the hippocampus, particularly subfield CA1, is an important communication hub not just at the local level, but in the large-scale connectome as well. These results showcase an important principle - the functional capacity of a given region or subnetwork cannot be fully discerned by analyzing the static structural connectivity of the brain. It is the communication between regions that engenders complex phenomena such as perception, cognition and action. Communication dynamics are the link between structure and function and thus represent a novel and important attribute of brain networks.
Chapter 6

General discussion

Recent advances in the quantitative analysis of large-scale brain networks suggest that system-wide communication may, in the future, help to understand why various aspects of neural function, such as attention and working memory, are limited. The studies presented in this thesis were designed to establish a framework for studying neural activity and network organization from the perspective of communication. The primary goal was to operationalize various aspects of communication and to show that these new measures can be used in a meaningful way to gain novel insights into brain structure and function.

Studies 1 and 2 described a method that uses functional data to characterize how information is assimilated and integrated by individual neural elements. Study 1 showed that the method can be used to capture consistent Gamma-like inter-event time distributions, as well as subtle individual differences that relate to how information is integrated. Study 2 showed a potential application of this method, whereby the face inversion effect could be explained in terms of disruptions to integrative processing. Studies 3 and 4 described a modeling method to characterize communication efficiency in large-scale networks. Study 3 showed that the rich club of hub nodes is an important topological feature that governs how signal traffic is distributed in the brain. In particular, the rich club serves to establish systematic deviations in information flow, such that some areas receive more and some areas less traffic. Study 4 showed that, despite its modest connectivity, CA1 is an important node in the brain, potentially acting as a convergence zone for signal traffic by virtue of its embedding the global topology.

The specific implications of each study are discussed in the individual chapters.
In the following section, I discuss how these findings can be placed in the context of recent trends in the field, and how they can be thought of from the perspective of several well-known principles.

6.1 A telecommunication framework

The methods described in this thesis represent an effort to advance beyond notions of activation and connectivity, which have traditionally dominated the neuroimaging literature. In activation studies, increases and decreases in mean activity, or in the amplitude or latency of evoked responses, are used to infer whether a particular neural element is specialized for some function. Typically, there is no underlying model of what operations neural activity represents. Likewise, functional connectivity studies focus on whether the activity of one neural element is statistically associated with the activity of another element. Effective connectivity studies are further able to resolve the directionality of influence for such relationships, allowing us to draw conclusions about whether changes in the activity of one neural element cause changes in the activity of another. Again, there exists no underlying model of how information is transformed and ultimately communicated.

At the same time, the brain is capable of constructing multimodal representations of the outside world, planning and executing responses, learning associations and contingencies on multiple time scales and inferring the intentions of others. These complex mental operations are unlikely to be the domain of some specialized region or pathway, but rather due to synergetic action of multiple distributed networks. The studies presented here are part of a growing trend in the community to consider how information is communicated in brain networks and a growing recognition that such a question requires one to go beyond the biophysical variables, such as cortical potentials and the BOLD response, and to use explicit models of what neural dynamics represent (Graham and Rockmore 2010; Goñi et al. 2013; da Fontoura Costa et al. 2007; Dimitriadis et al. 2012).

Graham and Rockmore (2011) have argued that, throughout the history of neuroscience, the brain has always been viewed through the prism of the current technological zeitgeist. Thus, for much of the twentieth century, the computer metaphor of brain function has been influential, with the brain conceptualized as a set of circuits
with specialized components, and neural function as a set of “programs” that run on that hardware. However, the Internet may be a more appropriate metaphor (Graham and Rockmore 2011). According to this view, the structural connectivity of the brain corresponds to the Internet, while the functions and functional networks enabled by this connectivity correspond to the World Wide Web.

The work presented in this thesis did not pursue the Internet metaphor per se, but was inspired by the similarities between brain networks and telecommunication systems and what those similarities suggest about how information flow is governed in brain networks. Like the Internet, brain networks are often characterized as being scale-free - they have hierarchically self-similar connectivity patterns, resulting in power-law degree distributions (Eguiluz et al. 2005)(but see also Achard et al. (2006); Modha and Singh (2010)). This suggests that communication strategies that can take advantage of the Internet topology may also be particularly effective for brain networks. Moreover, the Internet is organized around a central sub-network, or backbone, that carries a disproportionately high volume of signal traffic and serves to bind the network together. For instance, the National Science Foundation Network (NSFNET) was originally designed to connect NSF-funded supercomputing centers and has gradually evolved into a central sub-graph for Internet traffic (Mills and Braun 1987). This is remarkably similar to the putative rich club and its hypothesized role in anatomical brain networks.

Altogether, the architectural similarities between the Internet and brain networks suggest functional similarities as well. How information is projected, transformed and integrated by large-scale brain networks is surely one of the most important questions in human brain mapping. Thus, the similarities between the brain and various types of telecommunication systems may prove instructive. The results presented in this thesis offer a first look into the types of insights that may be possible by adopting a telecommunication-based framework.

6.2 Nonstationarity and neuroergonomics

The inter-event time method described in Studies 1 and 2 falls within a much broader literature concerned with the nonstationary nature of brain signals. A stationary stochastic process is one whose joint probability distribution does not change in time.
Despite the fact that brain signals have long been recognized to be nonstationary, there has traditionally been a focus on the mean signal with little regard for other moments and parameters.

Recent studies have increasingly focused on the metastable dynamics of neural activity. For example, the power spectra of electromagnetic neural activity are known to be highly time-dependent and characterized by transient episodes of synchrony and desynchrony (Pfurtscheller and Lopes da Silva 1999). Time-delay embedding methods (Stam 2005; McIntosh et al. 2008; Mišić et al. 2010; Vakorin et al. 2011a,b) as well standard deviations (Garrett et al. 2010, 2011, 2012) have been used to quantify the variability of neural activity and have revealed a rich dynamic range of brain signals that is systematically affected by a variety of factors, including development, aging, cognitive engagement and disease. Other methods have sought to segregate neural activity into states and have shown that neural activity is subject to frequent and often spontaneous state transitions (Freyer et al. 2009, 2012; Pascual-Marqui et al. 1995; Vakorin et al. 2013) that may reflect a tendency for dynamic functional reconfiguration.

The inter-event time method takes this philosophy further, by treating inter-event times as a stochastic output process. By fitting a distribution to this variable, the dimensionality of the data is reduced and the process can be fully summarized by a small number of parameters. Putative changes in how information is assimilated manifest as a nonstationarity of inter-event times. This nonstationarity is indexed by task-dependent changes in the parameters of the inter-event time distribution. Therefore, in this approach the nonstationary nature of brain signals across tasks is explicitly measured and used to make inferences about changes in information processing in those tasks.

Similarly, the queueing network method described in Studies 3 and 4 falls within the emerging domain of large-scale computational models. The shared characteristic of these models is that they are founded on realistic, whole-brain anatomical connectivity. Their goal is to investigate global network interactions by simulating local dynamics. The models only differ in what aspect of neural activity they represent. In linear dynamical systems models the dynamics are driven by local Gaussian noise and by inputs specified by the structural connectivity (Zamora- López et al. 2010; Sporns et al. 2000). Thus, these models are essentially filters. Nonlinear dynamical systems models, which are described in more detail in the General Introduction (Chapter 1), model the biophysics of populations of neurons. Here, local dynamics spontaneously
organize into ensembles to produce oscillatory activity at the global level. The collective oscillations effectively “enslave” the dynamics at individual nodes. Finally, stochastic models represent the flow of information in terms of random walks, where nodes of the network represent the states of a Markov chain (da Fontoura Costa et al. 2007; da Fontoura Costa and Sporns 2006; Dimitriadis et al. 2012; Goñi et al. 2013).

The model presented in Studies 3 and 4 is a stochastic model, but what sets it apart from other models in that class is the fact that individual signal units have a source and destination and that their trajectories are tracked. In that sense, the queueing network model bears some relation to the family of models that are referred to as “neuroergonomic” (Parasuraman and Wilson 2008; Liu et al. 2012). The goal of this approach is to optimize brain function with respect to real-life settings, in order to establish the most efficient use of resources to perform some task. The present model does something similar, because it seeks to characterize the communication efficiency of a given network. Altogether, the linear, nonlinear, stochastic and neuroergonomic computational models provide complementary information about structure-function relationships in brain networks.

6.3 Segregation, integration and neural context

The two methods presented in this thesis - inter-event times and queueing networks - both make use of queueing theory to characterize information flow, but how do they relate to each other? Here it is instructive to consider the two methods from the perspective of segregation and integration. In practice, the inter-event times method is applied to individual neural elements and in that sense it primarily addresses the question of segregation. Conversely, the queueing network method investigates the synergetic, network-wide communication patterns in the whole network and therefore addresses the question of integration.

However, the key idea behind the notions of segregation and integration is that the two co-exist in a complex system such as the brain and the same is true of these methods. The inter-event times method has an element of integration, because it measures how information from multiple sources assimilates at a particular node. For instance, the right fusiform gyrus produces more sustained integrative epochs in response to upright faces (Study 2). Similarly, the queueing network method has an
element of segregation, because the local connectivity of a node has some bearing on its role in the network. For instance, much of the convergence of signal traffic at CA1 could be attributed to a degree imbalance at area TF (Study 4). Therefore these findings speak to both segregation and integration in brain networks and are salient examples of why this notion has been so prominent and enduring in the field.

The present results are an example of how the coexistence of segregation and integration can be meaningfully related to the notion of neural context - the idea that the local function of a particular neural element is inseparable from the wider, network-level role of that element (McIntosh 2000; McIntosh 1999; Bressler and McIntosh 2007). For instance, Study 3 shows that the rich club of hub nodes is the most important sub-network in the brain. On one hand, a node has to have a high degree to be part of this sub-network. On the other, the topology of the network is organized such that signal traffic is directed towards some nodes and away from others, resulting in significant, systematic deviations from what would be expected on the basis of degree alone. Therefore, the dominant information flow patterns in the network are determined in some measure by both the local properties of individual nodes as well their embedding in the global topology.

6.4 Caveats and limitations

Details about methodological caveats and specific assumptions are discussed at length in the individual chapters, so I will not belabour them here. However, the oft-quoted sentiment, “all models are wrong, but some are useful”, attributed to the statistician George Box (Box and Draper 1987), seems appropriate. Namely, it is obvious that the queueing theoretic framework is, strictly speaking, wrong. The exchange of information in the brain is unlikely to be mediated via discrete signal units. There is unlikely to be a queueing mechanism that buffers information at each particular node. The very notion of a large-scale brain graph with discrete nodes and edges, which constitutes the foundation for these studies, as well as many other paradigms in the field, is problematic given the lack of well-defined boundaries between brain areas. Altogether, the queueing theoretic framework can provide only limited insight into information processing in the brain, because the results must be qualified by the many assumptions that the model entails.
At the same time, the present set of models is both necessary and useful. As I have argued above, models that relate biophysical neural activity to information processing are necessary to test hypotheses in cognitive neuroscience. More importantly, the specific models described here are useful. For example, Study 1 shows that it is possible to extract a single-subject measure from resting-state EEG that is remarkably consistent across subjects. Study 2 shows that the face inversion effect is due to a fractionation of the integrative steps that support face perception. Study 3 shows that the connectivity of the brain is organized in a way that prioritizes speed of information processing. Study 4 shows that the hippocampus is a critical node in the large-scale connectome. Therefore, these methods offer a unique perspective on structure-function relationships and can be used to test specific, novel hypotheses about information processing in brain networks.

6.5 Future directions

For the methods described here, I foresee at least two future applications. First, the queueing framework will allow individual variation in anatomical connectivity to be assessed from a functional perspective. For example, models can be tailored to the connectivity of individual participants (determined using diffusion-weighted imaging) and used to simulate communication for those networks. Specific aspects of communication (e.g. speed, reliability, etc.) can then be related to individual differences in various cognitive faculties, such as attention and memory. Likewise, clinical populations with altered connectivity (as in autism and schizophrenia) can also be studied by comparing simulations based on intact versus damaged networks. Moreover, the functional consequences of lesions can be studied by removing nodes and/or edges to determine how the flow of information gets re-distributed following this anatomical change. This group of applications is analogous to the Virtual Brain framework, which works with a nonlinear dynamical model instead of a queueing network model (Jirsa et al. 2010; Ritter et al. 2013; Leon et al. 2013).

Second, the queueing framework can naturally be used to investigate how information is routed in brain networks. In the present set of studies information flow was assumed to be diffusive. However, this may not be the case in real brain networks; the extent to which signal units are aware of global topology, and the strategies that
are used to take advantage of that topology, are open questions. A systematic test of how well different routing strategies fit real brain data could be carried out by assuming that patterns of empirical functional connectivity (derived from fMRI, for example), represent the dominant patterns of information flow in the network. Thus, the total throughput of signal units along individual projections could be treated as a measure of functional connectivity and compared to empirical functional connectivity for different routing strategies, to determine which strategy best reproduces empirical functional connectivity.

6.6 Summary

As the focus of neuroscience shifts from studying single regions to networks of regions, there is a need to articulate information flow and communication in the context of brain networks. A basic problem in modern cognitive neuroscience is that we are trying to infer cognition and information processing from neuroimaging measures that are inherently biophysical, leaving a gap in the transition from neurobiology to cognitive processes. There is a need for methods and analyses that embody some model of how information processing relates to the observable neural activity. In the present set of studies I have described two methods based on queueing analysis that capture (a) how information is assimilated by individual brain regions and (b) how system-wide communication depends on network topology. These methods offer novel and unique insight into diverse questions about neural structure and function.
Appendix A

Analytical model
In the following section I seek to confirm the numerical model presented in Chapter 4 via analysis. I attempt to derive an analytical queueing network model comprised of $N$ nodes. The network is represented as a binary adjacency matrix $R$ of size $N \times N$. Each element $R_{i,j}$ is equal to 1 if there is a link from node $i$ to node $j$. If there is no link from node $i$ to node $j$ then $R_{i,j} = 0$.

Each node is modeled as server that has to process a received signal unit before it is forwarded to a neighboring node. The routing of signal units is random, i.e. the signal units are forwarded to randomly selected neighbours.

The external arrivals to each node are modeled as a Poisson process. We assume that the sum of processing time at the intermediate node $j$ and its forwarding time is exponentially distributed with parameter $\mu_j$. In addition to biological considerations, we choose the exponential distribution because allows for a tractable analytical model and because it has a coefficient of variation equal to 1. The relatively conservative coefficient of variation allows for tighter performance bounds for many other sub-exponential service and forwarding times. We also assume that each node has a finite buffer and uses an impartial Last Come First Served (LCFS) service discipline with pushout. Thus, in the case of buffer overflow due to high input load, the signal unit that arrived least recently will be lost.

### A.1 Modeling a node

In the classical $M/M/1/K$ First Come First Served (FCFS) system with finite capacity, each arriving signal that finds the system (buffer and server) full is blocked and lost (Kleinrock 1972; Takagi 1993b). Alternatively, in the pushout model, the arriving signal is always accepted, while the signal that has been waiting for the longest time is pushed out of the system. In this non-preemptive system, a signal that is already in service cannot be pushed out. It can be shown that the blocking probability of the wider class of $M/G/1/K$ FCFS systems is equal to the pushout probability of a pushout system (Takagi 1993b). We will denote this value as $P_B$.

In addition, the probability distribution of buffer occupancy is the same for blocking and pushout systems as long as the service policy is impartial. However, it is also known (Takagi 1993b) that the mean waiting time of a signal that is eventually served in LCFS pushout system is smaller than the waiting time in a FCFS system (although
its variance is larger) (Takagi 1993b).

Let us denote total arrival rate to node \( j \) as \( \lambda_{in,j} \). Then offered load for node is denoted as:

\[
\rho_{o,j} = \frac{\lambda_{in,j}}{\mu_j}.
\]

(A.1)

Then, probability of having \( 0 \leq k \leq K \) messages in the node’s \( j \) buffer is given by:

\[
p_{j,k} = \rho_{o,j}^k \frac{1 - \rho_{o,j}}{1 - \rho_{o,j}^{K+1}}.
\]

(A.2)

Pushout, i.e. blocking probability is then equal to:

\[
P_{j,B} = P_{j,K} = \rho_{o,j}^K \frac{1 - \rho_{o,j}}{1 - \rho_{o,j}^{K+1}}.
\]

(A.3)

### A.2 Modeling a network of nodes

According to Burke’s theorem, the output of a M/M/m system is also a Poisson process (Kleinrock 1972). This result can be extended to finite M/M/m/K systems due to the memoryless property of the distribution of signal interarrival times, which is exponential. This result allows for the analysis of any feedforward interconnection of nodes using node-by-node decomposition analysis.

Likewise, Jackson’s theorem extends the previous result to a network of \( N \) nodes with feedback interconnections and allows a node-by-node decomposition (Kleinrock 1972). This fits our analysis since the queuing network based on the macaque anatomy has a significant feedback structure.

Let us define \( P_{dest} \) as the probability that the node that receives the signal unit is a destination for that unit. We also define the probability that the node that receives the unit has to forward it further as \( P_{tr} = 1 - P_{dest} \), i.e. the transit probability. Under the random routing assumption, these probabilities depend on the topology of the graph. A conservative estimate would be \( P_{tr} = \frac{1}{N-1} \) and \( P_{dest} = 1 - \frac{1}{N-1} \), although we will discuss less conservative approaches as well.

Let us assume that each node \( j \) in the network has an external input Poisson rate \( \gamma_j \). The remaining part of the input rate comes from nodes that project to node \( j \). By Burke’s Theorem, the output process of each node is also Poisson (Kleinrock 1972).

Under the decomposition approach, the input Poisson rate to node \( j \) is equal to:

\[
\lambda_{in,j} = \gamma_j + \sum_{i=1}^{N} \lambda_{out,i} r_{i,j}
\]

(A.4)
where \( r_{i,j} \) denote routing probabilities from node \( i \) towards node \( j \). Under random routing and a known connectivity matrix \( R \), routing probabilities are equal to:

\[
r_{i,j} = \frac{R_{i,j}}{\text{od}_i}, \tag{A.5}
\]

where \( \text{od}_i = \sum_{n=1}^{N} R_{i,n} \) is the output degree of node \( i \).

The output rate from each node is affected by blocking caused by the node’s finite buffer and by the portion of signals for which this node is a destination. Therefore, the output rate from node \( j \) is equal to:

\[
\lambda_{\text{out},j} = \lambda_{\text{in},j}(1 - P_{j,B})P_{tr} \tag{A.6}
\]

note that the output rate is also known as the throughput of the node. The carried load for node \( j \) can also be obtained as \( \rho_{c,j} = \rho_{o,j}(1 - P_{j,B})P_{tr} \).

Equations (A.3), (A.4) and (A.6) for each node form a system of \( 3N \) equations that can be solved in an iterative way. The first iteration of the system is started with only external input to the nodes and iterations are stopped when the relative increment of blocking probabilities between two iterations falls below a threshold of 0.0001.

### A.3 Modeling delay in a node

Let us first consider the FCFS system and the waiting time of the signal that arrives to the buffer and finds \( k-1 > 0 \) waiting messages and one signal being serviced. That signal has to wait for remaining service time of the signal being served plus \( k-1 \) complete service times.

Assume that the probability density function (pdf) of the message service time is denoted as \( b_j(x) = \mu_j e^{-\mu_j x} \) and its probability distribution function (PDF) is \( B_j(x) = \int_{y=0}^{x} b_j(y)dy \). The mean value of service time is \( \bar{b} = \int_{x=0}^{\infty} x b(x)dx \).

The Laplace-Stieltjes Transform (LST) of message service time in node \( j \) is denoted as:

\[
B_j^*(s) = \int_{0}^{\infty} e^{-sx} b_j(x)dx = \frac{\mu_j}{\mu_j + s}, \tag{A.7}
\]

The pdf of remaining service time is \( b_{j,+}(x) = \frac{1 - B_j(x)}{\bar{b}} \). The LST of the remaining signal service time is denoted as Cooper (1981); Takagi (1991a):

\[
B_j^{+,*}(s) = \int_{x=0}^{\infty} e^{-sx} b_{j,+}(x)dx = \frac{1 - B_j^*(s)}{s}. \tag{A.8}
\]
Due to the memoryless property, for the exponential distribution it holds that \( B^+(s) = B'(s) = \frac{\mu}{\mu + s} \).

Therefore, the LST of the delay in FCFS system with an exponentially distributed service time has the form:

\[
W_{fcfs,j}^*(s) = \frac{1}{1 - P_{jB}} \left( p_{j0} + \sum_{k=1}^{K-1} p_{jk} B^+_j(s) B^*_j(s)^{k-1} \right) = \frac{1}{1 - P_{jB}} \left( p_{j0} + \sum_{k=1}^{K-1} p_{jk} B'_j(s)^k \right)
\]

The mean value of the delay can be derived as \( W_{fcfs,j}^* = -\frac{dW_{fcfs,j}^*(s)}{ds} \bigg|_{s=0} \).

Analysis of message delay in the LCFS system with pushout is more involved due to the recursive nature of the delay process. In order to derive a distribution of this delay we first attain some intermediate results.

First, we define the LST of the joint distribution of service time and number of signals that arrive during service time as \( a^*_j(n) \). By the same token, we define the LST of the joint distribution of residual service time and number of signal arrivals as \( d^*_j(n) \). Expressions for \( a^*_j(n) \) and \( d^*_j(n) \) are presented as:

\[
a^*_j(n)(s) = \int_{x=0}^{\infty} \frac{\lambda_{in,j} x^n}{n!} e^{-(\lambda_{in,j} + n)x} b(x) dx d^*_j(n)(s) = \int_{x=0}^{\infty} \frac{\lambda_{in,j} x^n}{n!} e^{-(\lambda_{in,j} + n)x} b_j(x) dx
\]

(A.9)

Note that for the exponential distribution of service time \( a^*_j(n)(s) = d^*_j(n)(s) \). We also denote \( W_{jk}^*(s) \) as the LST of the delay for the waiting time of a signal that has \( k \) signals ahead at the end of service and is eventually served (index \( k \) ranges from \( 0 \leq k \leq K - 2 \)).

\[
W_{j0}^*(0) = 0 W_{jk}^*(s) = \sum_{n=0}^{K-k-1} a^*_j(n)(s) W_{jk+n-1}^*(s), \quad 1 \leq k \leq K - 2
\]

Using elimination, the system of equations (A.10) can be solved such that \( W_{jk}^*(s) = f(a^*_j(n)(s)), \quad n = 0..K - 2 \).

For signals that receive service, the LST of the delay has been presented as Takagi (1993b):

\[
W_{lcfs,j}^*(s) = \pi_{j0} + \sum_{k=0}^{K-2} a^*_j(k)(s) W_{jk}^*(s)
\]

(A.10)

where \( \pi_{j0} \) denotes the probability that the system is empty at signal departure time. Fortunately, for exponentially distributed service times \( \pi_{j0} = p_{j0} \).
Appendix A. Analytical model

The mean value of node delay can be obtained as

\[ \overline{W}_{lcfs,j} = - \frac{dW_{lcfs,j}(s)}{ds} \bigg|_{s=0}. \]

Finally, we need to mention that the signal response time for node \( j \) is the sum of signal waiting and processing times, i.e.

\[ T^{*}_{lcfs,j}(s) = W^{*}_{lcfs,j}(s) B^{*}(s) \quad \text{and} \quad T^{*}_{lcfs,j} = W^{*}_{lcfs,j} + \frac{1}{\mu}. \]

A.4 End to end delay (transit times)

Given the fact that CoCoMac graph is relatively large, well connected and that we use random routing, we used a computationally efficient approximation, whereby the number of traversed hops between source and destination node has a probability distribution similar to the geometric (Grimmett and Stirzaker 1992). However, a parameter of this distribution is the transit probability \( P_{tr} \), which can not be fully oblivious to the traversed path. If it was fully oblivious to the traversed path, then \( P_{tr} = \frac{1}{N-1} \), giving the Probability Generating Function (PGF) of the total number of hops:

\[ N_{hp}(z) = \sum_{k=0}^{\infty} P_{tr}^k (1 - P_{tr}) z^{k+1} \] \hspace{1cm} (A.11)

giving mean number of hops as \( \overline{N}_{hp} = \frac{dN_{hp}(z)}{dz} \bigg|_{z=1} = \frac{1}{1-P_{tr}} = N - 1. \)

A better approximation would assume that the transit probability decreases with each traversed hop until the number of hops reaches a number that is significantly larger than the diameter of the network. In the case of the macaque network with diameter equal to 4, we limit this to \( N - 1 \), which is 60 times larger than the diameter. In that case, for \( i \) traversed hops \( P_{tr,i} = \frac{1}{N-1} \), and the PGF for the number of traversed hops becomes:

\[ N_{hp,m}(z) = \sum_{k=0}^{N} \prod_{i=1}^{k} P_{tr,i} (1 - P_{tr,k+1}) z^{k+1} \] \hspace{1cm} (A.12)

The mean number of hops using this approximation is approximately \( N/2 \).

The LST of the end to end delay and processing time for signals originating in node \( j \) - according to both approaches - is:

\[ E^*_j(s) = \sum_{k=0}^{\infty} P_{tr}^k (1 - P_{tr}) T^{*}_{lcfs,j}(s)^{k+1} E_{lm}^*(s) = \sum_{k=0}^{\infty} \prod_{i=1}^{k} P_{tr,i} (1 - P_{tr,k+1}) T^{*}_{lcfs,j}(s)^{k+1} \] \hspace{1cm} (A.13)

with mean values of \( \overline{E}_j = (N - 1) \overline{T^{*}_{lcfs,j}} \) and \( \overline{E}_{lm} = - \frac{dE_{lm}^*(s)}{ds} \bigg|_{s=0}. \)
Additional averaging over all nodes in the graph gives \( \bar{E} = \sum_{j=1}^{N} \bar{E}_j / N \).

### A.5 Comparing analysis and simulation

As mutual verification of the analytic expressions derived above and the numerical simulations reported in Chapter 4, I compare the two for network- and node-level statistics (Figs. A.1,A.2). In general, the two methods are in good agreement for the node-level statistics (Fig. A.2). There is also considerable agreement for the network-level statistics (Fig. A.1) and the two methods show the same pattern of results. There are also some small differences between the two, particularly for transit times, and these are likely due to the fact that we had to approximate the distribution of “hops” from source to target node using the geometric distribution.
Figure A.1: **System statistics: analytical and numerical results.** The network-level results of the analytical model are shown against the numerical simulation, with 500 replications ($H = 20, \mu = 0.02$).
Figure A.2: **Node statistics: analytical and numerical results.** The node-level results of the analytical model are shown against the numerical simulation, with 500 simulations ($H = 20, \mu = 0.02$)
Appendix B

Exploring parameter space
In general, the model presented in Chapters 4 and 5 has two characteristic modes of operation. At low intensities (external arrival rates) the total number of signal units in the network fluctuates around some finite value and the system is stable. As the intensity is increased, there is a qualitative change in the system dynamics, characterized by a monotonic increase in the number of signal units in the network until all buffers are filled. This is analogous to a phase transition in dynamical systems theory and is sometimes referred to as a “jamming” transition Boccaletti et al. (2006); Tadić et al. (2004). The focus of the present studies was on the steady-state behavior of the network.

The key variable is the ratio between the arrival rate and service rate at each node. Therefore, I fixed the service rate ($\mu$) and varied the rate of external arrivals ($\lambda$). The latter effectively becomes an order parameter, capable of inducing a phase transition at some critical value.

The only other free parameter is the buffer capacity $H$. Buffer capacity is not a critical parameter in the sense that it cannot induce a phase transition in the system. Changes in buffer capacity will produce quantitative, but not qualitative, changes in system behaviour. To demonstrate this point, I show node-specific metrics at three different capacities in Fig. B.1. As buffer capacity is increased from $H = 5$ to $H = 100$, the total number of signal units in the system will increase because congested nodes can now hold more signal units in their buffers. Node contents are increased, particularly for the high-degree rich club nodes. Likewise, blocking probabilities remain unchanged, save for a few nodes where, due to the higher capacities, fewer signal units are lost. Node utilizations are slightly increased for all nodes, because fewer signal units are lost and there is a greater number of signal units in the system. However, the utilizations are fairly consistent across different buffer capacities, because the limiting factor in the utilization of each node is the ratio of arrival rate and service rate at each node.
Figure B.1: **Effect of buffer size.** The results of 500 simulations ($\lambda = 0.01$, $\mu = 0.02$), showing the utilization, blocking and node contents at each node for three different buffer sizes: $H = 5, 20, 100$. 
Appendix C

Similarity between network scenarios
Appendix C. Similarity between network scenarios

To assess the similarity between the patterns of results produced by the three different scenarios (CoCoMac, Small-World and Rich-Club), I employed the following procedure. For each network measure (e.g., transit time), I constructed a vector containing the results for the original, randomized and latticized networks at all simulation intensities (0.05, 0.1, 0.15, 0.2). For instance, for the CoCoMac (C) network and its randomized (R) and latticized (L) null models, the vector was organized as:

$$V_{\text{CoCoMac}} = \begin{pmatrix} C(0.05) \\ R(0.05) \\ L(0.05) \\ \vdots \\ C(0.20) \\ R(0.20) \\ L(0.20) \end{pmatrix}$$

(C.1)

This vector, representing the overall pattern of results for a given measure, was then correlated with analogous vectors constructed for the Small-World and Rich-Club scenarios. The correlation coefficients and associated p-values are displayed in Table C. Both the Small-World and Rich-Club scenarios are highly correlated with the CoCoMac pattern and this is due to the fact that increasing simulation intensity leads to similar behaviour in all networks, such as increasing throughput, blocking and utilization. Importantly, patterns associated with the CoCoMac scenario were consistently more highly correlated with the Rich-Club scenario ($\approx 0.99$) than with the Small-World scenario ($\approx 0.85$). Note that these results are invariant to whether the vector is constructed by arranging networks within intensities (as shown above) or intensities within networks.

Finally, to assess whether there is any difference between the correlation coefficients obtained for CoCoMac-Small World and CoCoMac-Rich Club scenarios, I applied Fisher’s r-to-z transformation. Following Cohen and Cohen (1975), the difference between the two transformed correlation coefficients is expressed as a z score (Table C, bottom row). The comparison is treated as a one-tailed test, and z scores greater than $|1.645|$ are taken to indicate that the CoCoMac-Rich Club correlation is significantly greater than the CoCoMac-Small World correlation.
## Appendix C. Similarity between network scenarios

<table>
<thead>
<tr>
<th></th>
<th>Transit Time</th>
<th>Throughput</th>
<th>Utilization</th>
<th>Blocking</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoCoMac vs Small-World</td>
<td>$r = 0.78$</td>
<td>$r = 0.88$</td>
<td>$r = 0.88$</td>
<td>$r = 0.86$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>CoCoMac vs Rich-Club</td>
<td>$r = 0.99$</td>
<td>$r = 0.98$</td>
<td>$r = 0.98$</td>
<td>$r = 0.99$</td>
</tr>
<tr>
<td></td>
<td>$p &lt;&lt; 0.001$</td>
<td>$p &lt;&lt; 0.001$</td>
<td>$p &lt;&lt; 0.001$</td>
<td>$p &lt;&lt; 0.001$</td>
</tr>
<tr>
<td>Fisher’s $r$-to-$z$</td>
<td>$z = 3.08$</td>
<td>$z = 2.14$</td>
<td>$z = 3.67$</td>
<td>$z = 2.27$</td>
</tr>
</tbody>
</table>

Table C.1: Comparing scenarios. To assess similarities in the patterns of results produced by the CoCoMac, Small-World and Rich-Club scenarios, we correlate the values of specific network metrics (transit time, throughput, utilization and blocking) across networks (original, randomized and latticized) and simulation intensities (0.05, 0.10, 0.15 and 0.20). Fisher’s $r$-to-$z$ expresses the difference between the CoCoMac-Small World and CoCoMac-Rich Club correlation coefficients as a $z$-score. Values greater than $|1.645|$ indicate that the CoCoMac-Rich Club correlation is significantly greater than the CoCoMac-Small World correlation.
Appendix D

Effect of resampling
In a discrete-event simulation, the system state is updated only when an event occurs. This event may be the creation of a signal unit, a signal unit entering a queue or leaving a server, or a signal unit being removed from the network. As a result, in many models system state is sampled at non-uniform time intervals. This is true of the model presented in Chapter 5 and 6, because system dynamics are governed by random variables, such as signal inter-arrival times and service times.

To facilitate conventional time series analysis, I used simple linear interpolation (also known as “table lookup”) to resample the system state at uniform time intervals. The new time series were effectively downsampled because an original simulation run had an average of 2.21 samples per time unit, while the interpolated time series were sampled once per time unit.

Fig. D.1 shows the effects of linear interpolation for one sample time series of network load. Compared to the original (Fig. D.1), the interpolated time series appears slightly smoother (Fig. D.1B), indicating some loss of high frequencies as expected. However, there appears to be little difference between the two time series.
Figure D.1: **Effect of interpolation.** Network load time series are shown for a single simulation ($\lambda = 0.01$, $H = 20$, $\mu = 0.02$). (A) Original time series, with non-uniform sampling. (B) Linearly interpolated time series, with uniform sampling. (C) Interpolated time series overlayed on the original time series.
Appendix E

Effect of rich club size
To demonstrate that network-level behavior reported in Chapter 4 could be replicated with a wide range of synthetic rich club networks, I created three additional such networks, with rich clubs of 10, 20 and 30 nodes. As described in the chapter, the networks were created from a random network by endowing a sub-set of nodes (the rich club) with greater connection density than the rest of the network, and an even greater connection density amongst each other.

Fig. E.1 shows the network-level metrics for each of these synthetic rich club networks, as well as their randomized and latticized versions. The pattern of results appears to be invariant to rich club size, and very similar to the network with 25 rich club nodes reported in the chapter. The small differences between the networks are likely due to the slight differences in density between the networks.
Figure E.1: **Effect of rich club size.** System statistics for 500 simulations ($H = 20$, $\mu = 0.02$) for three different synthetic “rich club” networks, with rich clubs comprised of 10, 20 and 30 nodes, out of 100 total nodes.
Appendix F

\textit{T}-values and \textit{p}-values
In Chapter 4 I report differences in network and node statistics for the CoCoMac network and its degree-matched randomized and latticized null models. Statistical assessment was performed by taking the average results for 100 simulations on the CoCoMac network and comparing these to 100 simulations on 100 realizations of the null networks. The average $t$-statistics and $p$-values associated with these comparisons are shown in Tables F.1 (transit times), F.2 (throughput) and F (node contents). Thus, Tables F.1 and F.2 correspond to Fig. 2 in the main manuscript, while Table F corresponds to Fig. 4 in the main manuscript.
### Appendix F. T-values and p-values

<table>
<thead>
<tr>
<th>intensity</th>
<th>C vs R</th>
<th>C vs L</th>
<th>R vs L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>$t = -3.14,$</td>
<td>$t = -10.05,$</td>
<td>$t = -9.16,$</td>
</tr>
<tr>
<td></td>
<td>$p = 2.2 \times 10^{-3}$</td>
<td>$p = 8.5 \times 10^{-17}$</td>
<td>$p = 3.8 \times 10^{-15}$</td>
</tr>
<tr>
<td>0.010</td>
<td>$t = -4.28,$</td>
<td>$t = -16.36,$</td>
<td>$t = -26.18,$</td>
</tr>
<tr>
<td></td>
<td>$p = 4.3 \times 10^{-5}$</td>
<td>$p = 6.8 \times 10^{-30}$</td>
<td>$p = 2.7 \times 10^{-46}$</td>
</tr>
<tr>
<td>0.015</td>
<td>$t = -3.76,$</td>
<td>$t = -11.16,$</td>
<td>$t = -14.33,$</td>
</tr>
<tr>
<td></td>
<td>$p = 2.9 \times 10^{-4}$</td>
<td>$p = 3.3 \times 10^{-19}$</td>
<td>$p = 7.0 \times 10^{-26}$</td>
</tr>
<tr>
<td>0.020</td>
<td>$t = -2.96,$</td>
<td>$t = -23.37,$</td>
<td>$t = -17.51,$</td>
</tr>
<tr>
<td></td>
<td>$p = 3.9 \times 10^{-3}$</td>
<td>$p = 4.4 \times 10^{-42}$</td>
<td>$p = 4.5 \times 10^{-32}$</td>
</tr>
</tbody>
</table>

Table F.1: Network comparisons for the Transit Time statistic. The average of 100 simulations on the CoCoMac (C) network was compared against 100 simulations on randomized (R) and latticized (L) null networks, for 100 null network realizations. The entries represent the average $t$-statistics and $p$-values for those 100 comparisons.

<table>
<thead>
<tr>
<th>intensity</th>
<th>C vs R</th>
<th>C vs L</th>
<th>R vs L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>$t = -2.81$</td>
<td>$t = -12.95$</td>
<td>$t = -11.63$</td>
</tr>
<tr>
<td></td>
<td>$p = 6.0 \times 10^{-3}$</td>
<td>$p = 5.1 \times 10^{-23}$</td>
<td>$p = 3.2 \times 10^{-20}$</td>
</tr>
<tr>
<td>0.010</td>
<td>$t = -3.07$</td>
<td>$t = -15.28$</td>
<td>$t = -17.30$</td>
</tr>
<tr>
<td></td>
<td>$p = 2.7 \times 10^{-3}$</td>
<td>$p = 8.5 \times 10^{-28}$</td>
<td>$p = 1.1 \times 10^{-31}$</td>
</tr>
<tr>
<td>0.015</td>
<td>$t = -2.57$</td>
<td>$t = -12.52$</td>
<td>$t = -16.74$</td>
</tr>
<tr>
<td></td>
<td>$p = 1.1 \times 10^{-2}$</td>
<td>$p = 4.0 \times 10^{-22}$</td>
<td>$p = 1.2 \times 10^{-30}$</td>
</tr>
<tr>
<td>0.020</td>
<td>$t = -2.66$</td>
<td>$t = -18.04$</td>
<td>$t = -11.30$</td>
</tr>
<tr>
<td></td>
<td>$p = 9.0 \times 10^{-3}$</td>
<td>$p = 4.6 \times 10^{-33}$</td>
<td>$p = 1.70 \times 10^{-19}$</td>
</tr>
</tbody>
</table>

Table F.2: Network comparisons for the Throughput statistic. The average of 100 simulations on the CoCoMac (C) network was compared against 100 simulations on randomized (R) and latticized (L) null networks, for 100 null network realizations. The entries represent the average $t$-statistics and $p$-values for those 100 comparisons.
### Table F.3: Comparisons for the node contents statistic

The average of 100 simulations on the CoCoMac network was compared against 100 simulations on randomized networks, for 100 null network realizations. The entries represent nodes with statistically significant differences, and the average $t$-statistics and $p$-values for those 100 comparisons.

<table>
<thead>
<tr>
<th>node</th>
<th>$t$-statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIT</td>
<td>$t = -10.95$</td>
<td>$p = 4.1 \times 10^{-22}$</td>
</tr>
<tr>
<td>CA1</td>
<td>$t = 34.25$</td>
<td>$p = 3.9 \times 10^{-85}$</td>
</tr>
<tr>
<td>23c</td>
<td>$t = 17.96$</td>
<td>$p = 1.9 \times 10^{-43}$</td>
</tr>
<tr>
<td>24</td>
<td>$t = -3.04$</td>
<td>$p = 2.7 \times 10^{-3}$</td>
</tr>
<tr>
<td>31</td>
<td>$t = 14.29$</td>
<td>$p = 2.7 \times 10^{-32}$</td>
</tr>
<tr>
<td>LIP</td>
<td>$t = -11.31$</td>
<td>$p = 3.2 \times 10^{-23}$</td>
</tr>
<tr>
<td>7b</td>
<td>$t = -5.92$</td>
<td>$p = 1.4 \times 10^{-8}$</td>
</tr>
<tr>
<td>8A</td>
<td>$t = -10.59$</td>
<td>$p = 4.6 \times 10^{-21}$</td>
</tr>
<tr>
<td>46</td>
<td>$t = -10.81$</td>
<td>$p = 1.1 \times 10^{-21}$</td>
</tr>
<tr>
<td>32</td>
<td>$t = 15.83$</td>
<td>$p = 5.2 \times 10^{-37}$</td>
</tr>
<tr>
<td>13a</td>
<td>$t = 23.73$</td>
<td>$p = 8.4 \times 10^{-60}$</td>
</tr>
</tbody>
</table>
Appendix G

Effect of simulation intensity
Figure G.1: **Effect of increasing simulation intensity.** Fluctuations in the total number of signal units present in the network during a single simulation run, shown for three different arrival rates (simulation intensities, $\lambda = 0.005, 0.01, 0.02$).
Appendix H

The rich club of the macaque brain
Figure H.1: **Rich club of the macaque network.** The spatial distribution of the rich club (RC), shown for two different rich club “levels” (adapted from Harriger et al. (2012)).
Appendix H. The rich club of the macaque brain

<table>
<thead>
<tr>
<th>RC1</th>
<th>RC2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>area</strong></td>
<td><strong>name</strong></td>
</tr>
<tr>
<td>13a</td>
<td>orbitofrontal area 13a</td>
</tr>
<tr>
<td>32</td>
<td>area 32</td>
</tr>
<tr>
<td>12o</td>
<td>orbital area 12</td>
</tr>
<tr>
<td>12l</td>
<td>medial area 12</td>
</tr>
<tr>
<td>11</td>
<td>area 11</td>
</tr>
<tr>
<td>9</td>
<td>area 9</td>
</tr>
<tr>
<td>46</td>
<td>cortical area 46</td>
</tr>
<tr>
<td>7b</td>
<td>area 7b</td>
</tr>
<tr>
<td>LIP</td>
<td>lateral intraparietal area</td>
</tr>
<tr>
<td>24</td>
<td>anterior cingulate</td>
</tr>
<tr>
<td>TH</td>
<td>temporal area TH</td>
</tr>
<tr>
<td>TH</td>
<td>temporal area TH</td>
</tr>
<tr>
<td>24c</td>
<td>rostral cingulate sulcus</td>
</tr>
<tr>
<td>23</td>
<td>area 23</td>
</tr>
<tr>
<td>25</td>
<td>area 25</td>
</tr>
<tr>
<td>Iai</td>
<td>intermediate agranular insular cortex</td>
</tr>
<tr>
<td>ENT</td>
<td>entorhinal cortex</td>
</tr>
<tr>
<td>PIT</td>
<td>posterior inferotemporal area</td>
</tr>
</tbody>
</table>

Table H.1: Rich club area names. The areas in the rich club (RC), along with their anatomical names. Note that RC1 is a subset of RC2, and RC2 includes all the areas listed, in both columns of the table.
Bibliography


