A recurrent dynamic model for correspondence-based face recognition

Philipp Wolfrum
Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt, Germany

Christian Wolff
Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt, Germany

Jörg Lücke
Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt, Germany, & Gatsby Computational Neuroscience Unit, University College London, UK

Christoph von der Malsburg
Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt, Germany

Our aim here is to create a fully neural, functionally competitive, and correspondence-based model for invariant face recognition. By recurrently integrating information about feature similarities, spatial feature relations, and facial structure stored in memory, the system evaluates face identity ("what"-information) and face position ("where"-information) using explicit representations for both. The network consists of three functional layers of processing, (1) an input layer for image representation, (2) a middle layer for recurrent information integration, and (3) a gallery layer for memory storage. Each layer consists of cortical columns as functional building blocks that are modeled in accordance with recent experimental findings. In numerical simulations we apply the system to standard benchmark databases for face recognition. We find that recognition rates of our biologically inspired approach lie in the same range as recognition rates of recent and purely functionally motivated systems.

Keywords: perceptual organization, object recognition, visual cognition, object representation, dynamic links, macrocolumns


Introduction

Over the past decades the task of visual object recognition has emerged as an intriguing and difficult scientific problem. Its many facets have been studied within and across disciplines such as physics, mathematics, biology, psychology, and computer science. Biological vision has in this context long been a source of inspiration for solutions to technical vision tasks because of its far-reaching capabilities that are unmatched by artificial systems so far. For instance, computational vision systems possess multiple layers of processing, which have been inspired by and are regarded as analogous to the multiple stages of processing in the visual system of animals and humans. In mammals the primary visual cortex (V1), the secondary visual cortex (V2), and the inferotemporal cortex (IT) constitute such stages. These different layers process and exchange information to infer knowledge from a given input. An important open question we address in this paper is how the information transfer between the stages of processing is organized. Models of visual object recognition can differ significantly in this respect. They range from systems with static connectivity for information transmission between layers (e.g., LeCun, Huang, & Bottou, 2004; Mel, 1997; Riesenhuber & Poggio, 1999; Rosenblatt, 1961), over systems that moderately manipulate transmission (e.g., Grimes & Rao, 2005; Walther, Itti, Riesenhuber, Poggio, & Koch, 2002), to systems that employ explicit mechanisms to dynamically modulate information transmission between any two stages (e.g., Grimes & Rao, 2005; Walther, Itti, Riesenhuber, Poggio, & Koch, 2002), to systems that employ explicit mechanisms to dynamically modulate information transmission between any two stages (e.g., Grimes & Rao, 2005; Walther, Itti, Riesenhuber, Poggio, & Koch, 2002). Likewise, models range from systems that process information in pure feedforward fashion (e.g., LeCun et al., 2004; Mel, 1997; Riesenhuber & Poggio, 1999; Rosenblatt, 1961) over systems with little or intermediate recurrence (e.g., Walther et al., 2002) to models with fully recurrent processing (e.g., Lücke et al., 2008; Wiskott & von der Malsburg, 1996). Systems with static connectivity for information transmission between processing stages are
usually feed-forward whereas systems with dynamic manipulation of information transmission are usually recurrent.

The great diversity of models of visual recognition is, at least partly, due to different strategies in addressing the problem of invariance—a problem that is central to object recognition in general. In one class of systems, here referred to as feature-based systems, invariance is achieved by letting signals from feature detectors at different positions (and scale and orientation) in an earlier layer converge onto one or a few units in a later layer (e.g., LeCun et al., 2004; Mel, 1997; Riesenhuber & Poggio, 1999; Rosenblatt, 1961; Walther et al., 2002). This strategy, often referred to as pooling, preserves feature information but not position information. Feature pooling is usually implemented in static architecture with an appropriate pooling operation such as selecting the maximal incoming signal (Riesenhuber & Poggio, 1999). Another class of systems, here referred to as correspondence-based (cf. Zhu & von der Malsburg, 2004), achieves invariance by actively routing signals through rapidly changing connections, transmitting information not only about the identity of features but also about their spatial arrangement, and mapping the structure of an object into an object-centered frame of reference (e.g., Anderson, Essen, & Olshausen, 2005; Arathorn, 2002; Hinton, 1981; Kree & Zippelius, 1988; Olshausen et al., 1993; Wiskott & von der Malsburg, 1996).

Feature-based systems and correspondence-based systems have been successful in different application domains (see Biederman & Kalocsai, 1997, for a discussion). Feature-based systems (e.g., LeCun et al., 2004; Riesenhuber & Poggio, 1999) are successful in classification tasks (but see Pinto, Cox, & DiCarlo, 2008, for a critical discussion). In these tasks the relative insensitivity of feature-based systems to small metric variation of object parts is advantageous. Also, the static connectivity in feature-based systems allows them to be tuned to specific image databases. A problem for these systems is, however, a strong sensitivity to background effects (as discussed, e.g., in Zhu & von der Malsburg, 2004, or Lücke et al., 2008), which often requires an additional segmentation mechanism.

Correspondence-based systems on the other hand prevail in recognition tasks in which small differences in features and their arrangements are important. A typical such task is face recognition. Face recognition has attracted much attention because it is at the focus of human attention and is of high commercial value. Furthermore, much is known about it in terms of psychophysics and neurophysiology. Compared to other tasks, an important point for the purposes of this paper is, in addition, the existence of stiff competitive tests on widely available image galleries (e.g., Messer et al., 2004; Phillips et al., 2005; Phillips, Moon, Rizvi, & Rauss, 2000). In the most comprehensive commercial test (Phillips, Grother, Micheals, Blackburn, & Tabassi, 2003, see also www.frvt.org), the best performing systems all used active mechanisms to establish correspondences between input and a memory.

Other than having different degrees of success in different application domains, some types of data about the mammalian visual systems seem to be explained better by feature-based systems whereas other data are more convincingly explained by correspondence-based models. A main argument for feature-based feed-forward recognition has, for instance, been the processing speed of the human visual cortex. Thorpe and coworkers have shown (Thorpe, 1988; Thorpe, Fize, & Marlot, 1996) that humans can decide whether an image contains an animal or not in less than 150 ms. In the area of face recognition, Debruille, Guillemin, and Renault (1998) found that event-related potentials (ERPs) in response to novel vs. known faces start to differ as early as 76 to 130 ms. Since such times are not much longer than the time required for a first wave of spikes to travel through the ventral stream after presentation of an image, it has been argued that visual recognition must be feed-forward. However, with closer inspection, such an interpretation seems to capture only part of the story. For instance, population codes can increase the speed of information transmission. The average spike rate of large excitatorily coupled neuron populations can be read out on a timescale that is much faster than the average spike latency of their single constituting neurons (van Vreeswijk & Sompolinsky, 1998). Thus, networks that have such “high gain” connectivity can respond very sensitively to subtle and fast input changes (similar to the principle of criticality; Bak, 1996). Our model makes use of such a population code (see A dynamic model of cortical columns section). Furthermore, independently of population coding, correspondence-based systems can react very fast if their dynamic connections have already been primed for a specific stimulus. This might be the case in simple classification tasks like in the experiments of Thorpe (1988).

A further argument in favor of correspondence-based mechanisms is psychophysical experiments that found priming or congruency effects to play a strong role (for a review, see Graf, 2006). Recognition performance and reaction times depend on the primed orientation (Jolicoeur, 1985; Lawson & Jolicoeur, 1999) or size of objects (Bundesen & Larsen, 1975). These findings show that it does take effort and time to align the external world with internal representations, suggesting active dynamic processes for correspondence finding rather than passive pooling operations. There is also abundant physiological evidence that connectivities in the visual system are not static. Shifting receptive fields have been found in lateral intraparietal cortex (Duhamel, Colby, & Goldberg, 1992; Kusunoki & Goldberg, 2003), in MT (Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006), and even in V2 and V4 (Luck, Chelazzi, Hillyard, & Desimone, 1997), suggesting that effective receptive fields in the visual
system change from one instance to the next to route and match the current stimulus of interest to representations in memory. Combining the evidence for feed-forward on the one hand and correspondence-based processes on the other it appears likely that the brain employs both strategies (as argued for example by Yuille & Kersten, 2006). This could be in the form of a first feed-forward sweep, which is fast and unconscious, and a second step of in-depth recurrent processing, which leads to conscious perception (compare Lamme, 2003). This view is consistent with findings by Johnson and Olshausen (2003), who report two ERP signals related to object recognition, an early presentation-locked one and a later signal that correlates in timing with the response times for recognition.

In summary, we have discussed three sources of evidence that argue for the existence of correspondence-based mechanisms:

1. the necessity to detect delicate differences between features and their arrangements in functional applications,
2. psychophysical findings suggesting active coordinate transformations taking place during visual information processing, and
3. evidence from physiological data about the mammalian visual system showing quickly changing neuronal receptive fields.

To help integrating these research directions we propose in this paper a correspondence-based model that is both neurally inspired and functionally competitive. Our model is similar to neural network approaches to correspondence-based recognition such as shifter circuits (Anderson et al., 2005; Olshausen et al., 1993) and dynamic link matching (Lades et al., 1993; Wiskott & von der Malsburg, 1996; Würtz, 1995). The neural dynamics used in our system is similar to the one recently suggested in Lüke et al. (2008)—a system that advances earlier approaches in that it computes feature similarities neurally and is fast in terms of recognition times. The system in Lüke et al. (2008) is only applicable to the correspondence problem, however, whereas the system discussed in this paper is capable of full recognition.

Our paper is organized as follows. In the Correspondence-based recognition section we briefly review the correspondence problem, discuss different neurally inspired approaches to correspondence-based recognition, and present the basic architecture of our model. In A dynamic model of cortical columns section we define and discuss the basic computational element of our system and in The network model section their connectivity and dynamic interaction. The model is simulated and applied to face-recognition databases in the Simulations section. The Discussion section discusses the system and the simulation results in the context of experimental findings and other systems in the literature.

Correspondence-based recognition

Central to correspondence-based systems is the so-called correspondence problem, which is illustrated in Figure 1. Figure 1A contains two stick figures as images, one on the input side and one on the model side. Both of these images are represented by a layer of feature detectors or nodes (black circles). The correspondence problem is now simply the problem of finding the connections between input and model nodes that link corresponding object parts, e.g., connections between head and head, neck and neck, etc. In Figure 1A the black connections or links show the correct correspondences as a subset of all possible ones.

As a prerequisite for correspondence finding, feature similarities must be computed. Unfortunately, in realistic applications high feature similarity is not sufficient to find correct correspondences. Different images of the same

Figure 1. The visual correspondence problem is the task of linking corresponding points between two images. (A) Input and model images are represented by arrays of feature nodes (black circles). All potential correspondences are symbolized by lines between the feature nodes. The correct correspondences are indicated as black lines. (B) Examples of wrong correspondences, although connecting equal feature types.
object may vary greatly, leading to high similarity between non-corresponding points (see, e.g., Wiskott, 1999). Figure 1B shows this in cartoon form, black lines connect the features with highest similarity, which results in wrong correspondences in this case. For realistic inputs such situations are very frequent and the ambiguities increase the more feature detectors are used. For a human observer, in distinction, it is easy to find correct correspondences, also in Figure 1B. The reason for this is that an object is defined by its features and their spatial arrangement. Correspondence-based systems therefore have to take both of these cues into account. Shifter circuits (Anderson et al., 2005; Olshausen et al., 1993) are relatively rigid in activating topologically consistent sets of links, whereas dynamic link matching (Lades et al., 1993; Wiskott & von der Malsburg, 1996; Würtz, 1995) has a more flexible dynamic control that lets neighboring links communicate directly with each other. In this paper, we use a flexible dynamics, as in earlier DLM systems, together with explicit units that control the connectivity between layers similar to control units in shifter circuits. A two-layer model of this type was suggested in Lücke et al. (2008) to study fast and neurally plausible solutions to the correspondence problem. Here, we study a three-layer system and its application to the more complex task of recognition.

The principal architecture of the system discussed in this work is shown in Figure 2. It consists of three layers: an Input Layer for image representation, an Assembly Layer, and a Gallery Layer as memory. The Assembly Layer establishes correspondences between input and memory. It recurrently integrates information about feature similarity, feature arrangement, and face identity. Given an input, the integration of these information components results in the system to converge to a state that represents a percept. Figure 2 sketches the system after such a convergence when it has correctly established correspondences between a person’s face stored in memory (i.e., in the Gallery Layer) and a given input image of this person. In the following the system’s architecture and neurodynamic mechanisms will be discussed in detail.

A dynamic model of cortical columns

The computational elements of our system are motivated by anatomical and physiological properties of the cortex on the scale of a few hundred microns. In particular, our modeling reflects the cortex’s minicolumn organization (see, e.g., Mountcastle, 1997) and the concept of canonical cortical microcircuits1 as, e.g., suggested by Douglas, Martin, and Witteridge (1989). That is, we take cortical columns as basic computational elements of our network and assume that all columns perform similar stereotypical computations. Depending on the perspective or the cortical area a cortical column is commonly referred to as macrocolumn (Mountcastle, 1997), segre gate (Favorov & Diamond, 1990), hypercolumn (Hubel & Wiesel, 1977) or simply column (e.g., Yoshimura, Dantzker, & Callaway, 2005) and, for instance in primary visual cortex, comprises roughly all neurons that can be activated from one point in visual space.

The analysis of the fine structure within a column suggests disjunct populations of excitatory neurons as functional elements. Anatomically, axons and dendrites of pyramidal cells have been found to bundle together and to extend orthogonally to the pial surface through the cortical layers. All neurons that directly contribute to one such bundle form a thin columnar module of just a few tens of microns in diameter (Buxhoeveden & Casanova, 2002; Peters & Sethares, 1996; Peters & Yilmaz, 1993). Together with associated inhibitory neurons (see, e.g., DeFelipe, Hendry, Hashikawa, Molinari, & Jones, 1990; Peters & Sethares, 1997) such a module was termed minicolumn (see, e.g., Buxhoeveden & Casanova, 2002; Favorov & Kelly, 1994; Mountcastle, 1997, 2003; Peters & Sethares, 1996) and was suggested as the basic computational unit of cortical processing (but see Jones, 2000; Rockland & Ichinohe, 2004, for critical discussions). More recent evidence for disjunct functional units within a cortical column comes from experiments using focal uncaging of glutamate combined with intracellular recordings (Yoshimura et al., 2005). It was found that a column has a fine structure of functionally relatively disjunct populations of layer 2/3 pyramidal cells. The relation of these populations to the cortical minicolumn has yet to be clarified, however. The main potential difference is that the concept of a minicolumn requires neurons in a population to be spatially adjacent whereas
for neurons in the functional populations described in Yoshimura et al. (2005) this is not necessarily the case.

Independent of the spatial arrangement of a column’s functional subpopulations, there is little dispute about the existence of lateral coupling of such populations via a system of inhibitory neurons (Peters & Sethares, 1997; Yoshimura et al., 2005). For example, in Yoshimura et al. (2005), the excitatory populations of layer 2/3 have been found to receive common and population-unspecific input from inhibitory neurons of the same layer as well as from inhibitory neurons of layer 4 (see also Dantzker & Callaway, 2000).

We will define our dynamic model of a cortical column in accordance with these experimental findings. To be somewhat independent of different terminologies used in different communities, we will refer to the cortical column simply as column (instead of, e.g., macrocolumn or hypercolumn) and we will refer to its functional subpopulations as the column’s units.

In our model of a column we use dynamic variables that describe population rate activity following mean-field arguments discussed in numerous contributions (see, e.g., Gerstner, 2000; Latham, Richmond, Nelson, & Nirenberg, 2000; Lücke & von der Malsburg, 2004; Marti, Deco, Giudice, & Mattia, 2006; van Vreeswijk & Sompolinsky, 1998; Wilson & Cowan, 1973, for a columnar model). We describe the unit’s neural activity by a differential equation called modified evolution equation. This equation represents our model of inhibition among the column’s units and is a generalization of the well-known deterministic evolution equation (see, e.g., Eigen, 1971).

The activity $x_i$ of the $i$th unit in a column of $K$ units is given by

$$\tau \frac{d}{dt} x_i = x_i^\nu I_i - x_i \sum_{j=1}^{K} I_j x_j,$$  

where $\tau$ is a time constant and the exponent $\nu$ parameterizes the competition strength among the units. $I_i$ represents the input signal to unit $x_i$.

For $\nu = 0$, there is no competition, and Equation 1 simplifies to

$$\tau \frac{d}{dt} x_i = I_i - x_i \sum_{j=1}^{K} I_j x_j.$$  

In this case, all units represent their input proportionally, while the interaction term $\sum I_j x_j$ leads to activity normalization in the column (see Appendix A for a proof). For $\nu = 1$, on the other hand, we have strong competition among the units, leading to winner-take-all (WTA) behavior (see Appendix A).

In our model of object recognition we assume that there are two types of columns with different functions.

Dynamically, they only differ in the use of the competition parameter $\nu$:

1. **Feature columns** represent their input in a linear fashion. Consequently, the units in a feature column have no need to compete among each other, i.e., for them the parameter $\nu = 0$.

2. **Decision columns** show a WTA behavior leading toward a state where only the unit getting the strongest input remains active. These units receive a $\nu$ signal that linearly rises from 0 to 1.5 So they start out with linear dynamics like feature columns. With rising $\nu$, competition increases, finally leading to a WTA behavior that leaves only the unit with the strongest input active. The typical dynamics of a decision unit is shown in Figure 3.

The crucial computations will in our system be performed by decision columns, whereas feature columns serve for information representation. Both kinds of columns may actually have the same neural substrate with the only difference that feature columns do not receive (or just do not respond to) the $\nu$ signal.

In the networks that we will introduce in the following section, the units of a column communicate with units of other columns. For this communication, a column scales the output activities of its $K$ units such that its output
energy (i.e., the 2-norm of the column activity vector) stays constant\(^3\):

\[
\bar{x}_i := \frac{x_i}{\sqrt{\sum_{j=1}^{k} x_j^2}}.
\]

This kind of output normalization is advantageous for maintaining homeostasis in networks of columns and may be carried out by neurons in layer 5 of the cortex as suggested by Douglas and Martin (2004).

### The network model

In Lücke et al. (2008) a model for correspondence finding is described that makes use of a population code within cortical columns, which allows fast point-to-point matching between two patterns (estimated in the range or below 100 ms). Here we extend this model to a system that matches images of different geometry and can compare input images to a gallery of many stored models simultaneously. Preliminary results of this work have been described in Wolfrum, Lücke, & von der Malsburg (2008).

Our network is made up of **layers**, which loosely correspond to the different cortical areas that make up the visual system (we are not speaking here of the layers of anatomically different neurons that can be distinguished within one area of cortex). Layers are organized topologically, with a topology that may be stimulus space, like in V1 and somatosensory cortex, or a more abstract space. The layers of our network interact recurrently and activity collectively converges toward a final state that represents the “percept” of the network, in our case the possible recognition of a face.

Layers may contain both feature columns and decision columns. If we assume every feature column to represent all relevant features at one position of a retinal image, then layers of feature columns can represent whole images. The network introduced below uses layers of two different spatial arrangements:

1. **Rectangular grid.** Straightforward representation suitable for any image. Every column represents one specific geometric location (see Figure 4A).
2. **Face graph structure.** An arrangement specifically suited for faces, where each column represents an important landmark position on a face (see Figure 4B). Note that in this case, a column does not necessarily represent a fixed spatial location in the image, but rather a fixed semantic location (nose, mouth, eye, chin, etc.). Spatial locations of landmarks can change according to the face they represent.

The network consists of the following three layers (see Figure 5):

1. **Input Layer \( \mathcal{I} \):** Represents the input image in a rectangular grid.
model we use a predefined set of Gabor wavelets that appropriately sample orientation (over 8 orientations) and spatial frequency (over 5 scales) space, resulting in a number of $K = 40$ features at each point. That is, we use Gabor filter responses to model the RFs of the feature units in the Input Layer. For extracting the filter responses, we use the standard Gabor transform, as described in Appendix A. As feature values we use the magnitude $J$ of the responses, thus ignoring Gabor phase, to model complex cell responses (Hubel & Wiesel, 1977). Implicitly, Gabor phase is still represented by the positions of the feature columns in the input image. In applications using Gabor features it has turned out that with $K = 40$, as above, good results can be achieved (Wundrich, von der Malsburg, & Würtz, 2004). Performance increases for more wavelets, but 40 represents a good compromise between performance and computational cost.

Each Input Layer unit being responsive to a certain Gabor feature $J_i^p$ at its position $p$ on the input grid, the unit activities follow the dynamics (cf. Equation 2)

$$
\frac{d}{dt} x_i^p = J_i^p - \sum_{j=1}^{K} J_j^p x_j^p.
$$

### Assembly layer

The Assembly layer integrates intermediate information from both the input image (represented in the Input Assembly units) and the gallery (represented by the Gallery Assembly units, see Figure 6). The role of the Input Assembly is to represent a normalized version of the input image, while the Gallery Assembly accommodates a weighted average of all Gallery faces. This information is organized in a face graph arrangement with $Q = 48$ landmarks (see Figure 4B). Since the face graph in the Assembly Layer has to be able to represent many different faces, we determine its geometry by averaging over several hundred face graphs of individual faces.

The columns of the Input Assembly and Gallery Assembly are feature columns, i.e., they integrate their inputs (defined below) according to Equation 2. The input $I^j_i^A$ to the $i$th Assembly unit at position $q$ of the face graph is a weighted sum of the $i$th Gabor feature at all grid positions $p$ of the Input Layer, modulated by the respective control units:

$$
I^j_i^A = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} C_{p,q} I_{p,j},
$$

with $C_{p,q}$ being the output strength of the dynamic link (see below) controlling the flow of the output of Input column $I_p$ to Input Assembly column $I^j_i^A$.

The input $I^G_i^A$ to a Gallery Assembly unit at position $q$ is the superposition of all Gallery activities at the same positions.

---

**Figure 6.** Information flow in our system. Visual information in form of Gabor jets $J$ extracted from an input image activates the Input Layer $I$. It flows to the Assembly Layer (Input Assembly, $IA$, see Figure 6) and from there to the Gallery $G$, where it activates via receptive fields $v$ some memories more strongly than others. Information representing the active memories (stored in projection fields $w$ analogous to $v$) flows back to the Gallery Assembly $GA$, from the Input Layer to the Input Assembly is modulated by the control units $C$, which in turn are driven by the similarity of those image patches in the Input Layer and the Gallery Assembly that they connect. By activating those control units that connect positions of the Input Layer containing similar information as the Gallery Assembly, the system effectively focuses on those parts of the input image that contain visual information most similar to the current reconstruction in the Gallery Assembly, formed by superposition of active units in the Gallery Layer. The thick black arrows represent the competition among the decision columns of which the Gallery and the control columns consist. The symbols correspond to those used in the text.

2. Assembly Layer: Integrates intermediate information from both the input image (represented in the Input Assembly units $IA$, see Figure 6) and the gallery (represented by the Gallery Assembly units $GA$).

3. Gallery Layer $G$: Represents all gallery faces in terms of the weights of its afferent and efferent connections to the Assembly Layer.

The following three subsections describe these layers in detail.

**Input layer**

The Input Layer represents the input image using 400 feature columns arranged in a rectangular grid of $P = 20 \times 20$ points. Each feature column represents by its units’ activities $K$ features extracted from the image at that position.

If we neglect color and binocularity, the response properties of neurons in primary visual cortex are commonly described by the well-known Gabor wavelets (Daugman, 1980; Jones & Palmer, 1987; Ringach, 2002). Note that other wavelets like the Cauchy filter (Wallis, 2001) might presumably also lead to good results. In our
landmark, filtered/multiplied by the feature vector represented by that respective Gallery unit:

\[ p_{q,i}^{GA} = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} w_{q,m,i} g_{q,m}, \]  

(6)

with the “afferent weight” \( w_{q,m,i} \) representing the strength of Gabor feature \( i \) in landmark \( q \) of Gallery image \( m \) (of \( M \) in total).

**Control units**

The Assembly Layer also contains the control units mentioned above, which mediate the signal coming in from the Input Layer. These control units provide potential connections (dynamic links) between every Input Layer point to every point in the Input Assembly. The activity of the control units is driven by the feature similarity of the corresponding points in the Input Layer and the Gallery Assembly. That is, the similarity between the non-normalized input face in the Input Layer and the weighted average face in the Gallery Assembly controls via the control units how input information flows to the Input Assembly. In that sense the control units define a geometric mapping between Input and Assembly Layers. Additionally to the feature similarity input, control units get support from neighboring control units that represent similar mappings (see Figure 7 and paragraph below for details).

The dynamic links are decision units, meaning that their dynamics follow Equation 1. The input \( I_C \) to a dynamic link \( C_{p,q} \) connecting input position \( p \) and assembly position \( q \) is given by the scalar product between both column outputs plus a topological interaction term:

\[ I_C_{p,q} = \sum_{i=1}^{K} I_{p,i} g_{A,q,i} \]

\[ + \frac{c_{top,c}}{|\text{neighbors}|} \sum_{\rho,q} f_{top}(p,q,\rho,q) C_{\rho,q}, \]  

(7)

where \( c_{top,c} \) defines the maximal strength of topological interaction between control units (see below), and \( |\text{neighbors}| \) is the number of topologic neighbors the control column has in the face graph.

**Topological cooperation among control units**

As mentioned before, there is topological cooperation among the control units of the Assembly Layer. The purpose of this cooperation is to establish a continuous mapping between the different geometries of the Input Layer and the Input Assembly. A given dynamic link connects a specific column \( A \) of the Input Layer with a column \( B \) of the Input Assembly. Due to the geometry of both layers, both columns represent distinct positions \( z_A \) and \( z_B \) in retinal coordinates and internal image representation space, respectively. Consequently, the dynamic link between them represents a certain geometric distance \( d_i = z_B - z_A \).

The idea is now to have topological connections in order to support parallel or near-parallel dynamic links. Therefore we define the strength of a topological connection between any two dynamic links \( i \) and \( j \) whose columns are neighbors in the face graph through a monotonically decreasing function of their nonparallel-ity/disparity:

\[ f_{top}(i,j) = f(||d_j - d_i||_2). \]  

(8)

Here we use a linearly decreasing thresholded function of the form

\[ f(y) = \max \left( 0, 1 - \frac{y}{\beta} \right). \]  

(9)
Thus topological interaction is always positive and acts only between more or less (depending on $\beta$) parallel neighboring links. This principle is depicted in Figure 7. To obtain the topological interaction $f_{\text{top}}$ in Equation 7 between two control units $C_{p,q}$ and $C_{p,q}$, we first calculate from the coordinates of the columns they control in the Input and the Assembly Layer the geometric distances $d_{p,q}$ and $d_{p,q}$, represented by them. From these we calculate the disparity of the two control columns according to Equation 8 and the topological interaction via Equation 9.

### Gallery layer

The Gallery Layer represents all $M$ gallery face images in a face graph of $Q$ decision columns. Each column corresponds to one landmark, with the units representing specific feature vectors for the individual faces at the respective landmarks by their afferent and efferent connections (see Figure 6). The units in the Input Assembly activate the Gallery units through receptive fields $\nu$ representing the stored facial landmark features, activating more strongly units of faces that are similar to the normalized input image in the Input Assembly:

$$f^G_{q,m} = \sum_{i=1}^{K} v_{q,m,i} I_A_{q,i} + \frac{c_{\text{top},G}}{Q} \sum_{q=1}^{Q} G_{q,m}. \tag{10}$$

Additionally, there is interaction among the Gallery units, with $c_{\text{top},G}$ defining how strongly Gallery units representing the same face cooperate. That is, all landmarks that belong to the same face cooperate, and at each landmark the corresponding features of all different faces compete.

The Gallery projects a weighted superposition of its stored faces to the Gallery Assembly through efferent weights $w$ that are identical to its afferent weights $v$ (cf. Equation 6). Point-to-point comparison with the Input Assembly and competition among stored models leaves only the correctly recognized identity active in the end.

### Simulations

We now simulate the dynamics defined in the above sections using natural images of faces as input and as memories in the gallery. To integrate Equation 1, we simply use the Euler method but adapt its time step dynamically to the average change of activity in the network in order to keep the system stable. All units have a small, but non-zero initial activity $x(0) = 0.01$. The units in the Input Layer, which receive input directly from the incoming image (cf. Equation 4), quickly converge to a state where they represent the input image via the different Gabor feature values at all grid positions. This information flows to the Input Assembly modulated by the activities of the control units (Equation 5), which connect every point in the Input Layer with every point in the Input Assembly. Since initially all control units have equal activity, this leads to a superposition of image information from all Input Layer points at each Input Assembly location, resulting in a featureless, more or less homogeneous image in the Input Assembly (first image in Figure 8). In the Gallery Layer, all faces are equally active initially. The Gallery Assembly, which receives input from all Gallery units (Equation 6), will therefore initially receive a superposition of all Gallery faces, resembling an “average face” (like the first image in Figure 9).

To each control unit in the Assembly Layer a unique pair of feature columns is assigned, one in the Input Layer and the other one in the Gallery Assembly. The control units are driven by the similarity (expressed in terms of the scalar product) of the information stored in their dedicated feature columns, see Equation 7. Therefore control units that connect points of the average face with similar input points will become stronger, while control units representing irrelevant matches will be weakened. Over the process of recognition, the activity distribution of the control units becomes more and more sparse, until it finally represents a unique mapping between the Input Layer and the Assembly Layer (see left column in Figure 8).

Since purely local similarity of images can be quite ambiguous, the additional topological interaction among the control units is necessary in this process to achieve a globally consistent match. As the information flow from the Input Layer to the Input Assembly is modulated by the control units, the image in the Input Assembly will start to develop from a gray nondescript superposition to a more and more clear version of the input image (right column in Figure 8). It may be shifted and possibly distorted such that it conforms to the topology of the face graph of the Gallery Assembly.

The image information in the Input Assembly in turn acts as input to the Gallery units, where it gets filtered through the individual receptive fields of the units (Equation 10), exciting those units more that represent faces more similar to the input image. Due to competition between the units of each Gallery column and cooperation among units of different landmarks representing the same face the Gallery will start to favor some of the stored faces over others (cf. left column of Figure 9). This in turn changes the image in the Gallery Assembly from an average face to a superposition that is already biased toward one or several of the better fitting gallery faces (second and third face images of Figure 9). This sharpened target face now helps to position the normalized input image even more precisely, and so forth. In the final state, the Input Assembly will contain a shifted and maybe distorted version of the input image, while in the
Gallery Layer the units of only one face are still active, and the Gallery Assembly contains a copy of that face of the Gallery that the system judges to be most similar to the input image. Note that we can numerically simulate dynamics (Equation 1) without specifying a value of the time constant \( C \). As long as the simulation time \( T \) remains constant relative to \( C \), simulation results will be independent of \( C \). The question of how the time course of the dynamics translates to recognition times in biological terms does, however, crucially depend on the actual choice of \( \tau \). Using a time constant of, e.g., \( \tau = 0.2 \) ms results in a system that selects the winning subpopulations of its decision columns in about 80 ms (with \( T = 400 \tau \), compare Figure 3). The whole network could consequently converge to a face position and identity within about the same time. Numerical simulations of a single column with explicitly modeled spiking neurons suggest an even smaller time constant (see Lücke and von der Malsburg, 2004 or compare Muresan and Savin, 2007 for population activation times on the order of 10 ms, which suggest similarly fast deactivation times). For more detailed conclusions it would, however, be necessary to model a system with a setup like the presented one but based on detailed single neuron models instead of abstractions for populations. The velocity of recognition in such a system would ultimately depend on the time constants of the single neurons’ ion channels and dendritic and axonal conduction times in such a case.

To quantitatively compare our system to other approaches, we tested it on the FERET (Phillips, Wechsler, Huang, & Rauss, 1998) and the AR (Martinez & Benavente, 1998) benchmark databases. We followed the testing protocol of Phillips et al. (2005) and of Tan, Chen, Zhou, and Zhang (2005). The FERET database contains images of 1196 individuals, while the subsets of the AR database used by Tan et al. (2005) and by us contain 100 faces. The measured recognition rates (see Table 1) show that our system is competitive with purely functionally motivated approaches, although it cannot compete with the best performing system for each individual test category. It should be noted, however, that most of those systems were only tested on a single database, allowing
fine-tuning for these specific circumstances, while we used the same parameter settings on both databases. In general, our focus was rather on creating a neurally plausible system than on investing much effort in parameter tuning. For a more extensive discussion of these results see Wolfrum et al. (2008).

**Discussion**

We present here a fully neural model for face recognition that goes in essential ways beyond previous work in our own group (Wiskott, Fellous, Krüger, & von der Malsburg, 1997; Wiskott & von der Malsburg, 1996; Zhu & von der Malsburg, 2004) and the recent model in Lücke et al. (2008). It combines findings from psychophysics, imaging studies, and physiology, while performing competitively on benchmark tests for face recognition.

The basic building block of our system is a model of the cortical column, the biological relevance of which has been discussed earlier (Lücke & von der Malsburg, 2004). That model makes use of a population code for stimulus representation, thus allowing significantly faster computations than with rate codes of single neurons (for early
arguments in this direction see, e.g., van Vreeswijk & Sompolinsky, 1998). An essential ingredient of the model is formed by dynamic links. These are synaptic connections that are modulated by the activity of control units, whose activity in turn is controlled by signal comparisons. For a discussion of the hypothesis that control units might be formed by astrocytes (see Möller, Lücke, Zhu, Faustmann, & von der Malsburg, 2007). There is strong experimental evidence that receptive fields of neurons are not static (see Luck et al., 1997, and other references in the Introduction section), suggesting the existence of dynamic links. Other models in the literature argue for similar concepts, like the control units of Olshausen et al. (1993) or Sigma-Pi neurons (e.g., Weber & Wermt, 2007).

In anatomical terms, the different layers of our model can be interpreted as follows. The input layer represents incoming image information by Gabor wavelets, which resemble the receptive field properties in primary visual cortex (V1). The biological counterpart of our Assembly Layer would be an area like central or anterior inferotemporal cortex. Neurons here respond to stimuli from large parts of the visual field, and they code for complex temporotemporal cortex. Neurons here respond to stimuli from Layer would be an area like central or anterior inferotemporal cortex. Neurons here respond to stimuli from large parts of the visual field, and they code for complex

can be interpreted as follows. The input layer represents incoming image information by Gabor wavelets, which resemble the receptive field properties in primary visual cortex (V1). The biological counterpart of our Assembly Layer would be an area like central or anterior inferotemporal cortex. Neurons here respond to stimuli from large parts of the visual field, and they code for complex shapes (Tanaka, 1996, 2003) similar to the face parts represented by the Assembly Layer. The fact that information about object position and scale can be read out from IT neurons (Hung, Kreiman, Poggio, & DiCarlo, 2005), which disagrees with the assumptions made by pure pooling models, points to the possibility of our control units residing there as well. Of course, in the cortex the mapping from V1 to IT does not happen directly, but via intermediate stages including V2 and V4. This is not accounted for in our current model but will be included in future extensions. We have described previously the likely form (Wolfrum & von der Malsburg, 2007b) of such routing over several stages and a possible ontogenetic mechanism (Wolfrum & von der Malsburg, 2007a). Finally, the Gallery of our model might correspond to an area like the fusiform face area (ffa), which is specialized for face recognition (Kanwisher & Yovel, 2006; Tsao, Freiwald, Tootell, & Livingstone, 2006). Note that the detection of faces (as distinct from recognition) is not modeled by us. Also in the brain, this appears to happen outside of ffa. Summerfield et al. (2006) find neurons in medial frontal cortex that are selectively active when subjects have to make a face vs. non-face decision, independently of face identity. Likewise, prosopagnosia patients recognize objects as faces but cannot identify them (Zhao, Chellappa, Phillips, & Rosenfeld, 2003). As discussed before, face recognition is special because faces have a generic shape, but recognition from thousands of individuals requires high sensitivity to detailed differences. This might become possible through competitive interaction (which in fact is the mechanism by which recognition happens in our Gallery Layer) in the small and compact ffa (Kanwisher, 2006). Apart from faces, there is evidence suggesting that ffa can also serve as an area of expertise for other object classes (Gauthier, Skudlarski, Gore, & Anderson, 2000; Tarr & Gauthier, 2000). In the same sense, our model is not confined to face recognition but could be used for recognition of any kind of object type that has a prototypical shape and requires high sensitivity to small differences among objects.

Our system is correspondence based, as in Arathorn (2002), Hinton (1981), and Olshausen et al. (1993), in distinction to the majority of object recognition models (e.g., Mel, 1997; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007). As discussed in the Introduction section, we offer here several arguments in favor of correspondence-based object recognition. One is functionality—while feature-based systems are very good at classifying images into categories (Fei-Fei, Fergus, & Perona, 2006; Serre et al., 2007), they perform poorly in areas like face recognition, where high sensitivity to subtle image differences is required. As pointed out in the Introduction section, according to the latest test (Phillips et al., 2003, see also www.frvt.org) the best-performing commercial systems are all correspondence based.

Second, the system is extensible to active perception and vision in dynamic environments. In distinction to feed-forward models without dynamics (like Serre et al., 2007), which behave like static filter banks, in the proposed system the activity traces of previous processing (a previous $\nu$ cycle) can influence the outcome of current perception. This permits to model temporal effects in perception. Instances of these are priming or congruency effects, which abound in psychophysics (for a review, see Graf, 2006) and which can be easily replicated in our framework (Wolfrum & von der Malsburg, 2008): preactivating a subset of the control units at the beginning of a $\nu$ cycle (equivalent to spatial attention) results in a bias in favor of a specific location. For a large input image containing several faces, the system preferentially processes and recognizes objects at that position. Note that the priming of pathways instead of feature activity merely biases the system toward a certain decision without distorting the content that is being processed. In contrast, for example the model of Deco and Rolls (2004) implements spatial attention by increasing activity at the input level, which might lead to a distortion of information content. Also experimental findings (Luck et al., 1997) suggest that priming shifts effective receptive fields (that is, pathways) without changing activity within the afferent layer. Similarly, the priming of objects is possible with the help of preactivating search images in the form of arbitrary combinations of facial features in the Gallery Layer, leading to preferential detection and recognition of a similar face among several others in the visual field. In distinction, a feed-forward model would only permit the preactivation of objects for which there is already a dedicated representation. Furthermore, preactivation of dedicated, implicit representations would not generate an explicit search image at lower layers but a large unstructured activation of all features that potentially could give rise to the primed object.
A third argument for our model and for correspondence-based recognition in general is its use of explicit object representations. In a cardinal cell, whose activity stands for a specific object or property, representation is implicit, as the activity of the cell does not express the structure of the object, which is only implicit in the synaptic patterns that define its firing condition. An explicit representation, on the other hand, makes a wealth of distinctions available to recipients elsewhere in the brain. An explicit representation therefore forms not just a sample point in a structure space but represents a whole space of variations. However, in distinction to the generality of the retinal representation, which captures only properties common to all visual scenes, the ultimate goal of the visual system is to lead to explicit representations that are tightly committed to distinctions applying to the actually inspected scene and that permit strong probabilistic coupling to variables in other modalities, such as action planning, motion control, or language. It is these explicit, temporarily committed representations that give us the feeling of being in direct contact with the visual reality out there.

How does our model reflect these distinct types of representation? The representation in the Input Layer is explicit and uncommitted, and its generality is restricted only by the incomplete sampling of Gabor space. The representation in individual units in the Gallery Layer is implicit, highly specific and completely committed (to individual landmarks in individual faces). Given its activity state, the Gallery Layer creates, via its output connections to the Gallery Assembly, an explicit representation of a face. As long as inhibition is low, this representation is still uncommitted to an individual (see Figure 9, uppermost image), whereas at the end of the selection process it is fully explicit and committed to one individual face. While models of object recognition that explicitly reconstruct the input are common in probabilistic modeling (conceptually discussed, e.g., in Yuille & von der Malsburg, 2008) as well as other image representations, which only implicit in the synaptic patterns that define its firing condition. An explicit representation, on the other hand, makes a wealth of distinctions available to recipients elsewhere in the brain. An explicit representation therefore forms not just a sample point in a structure space but represents a whole space of variations. However, in distinction to the generality of the retinal representation, which captures only properties common to all visual scenes, the ultimate goal of the visual system is to lead to explicit representations that are tightly committed to distinctions applying to the actually inspected scene and that permit strong probabilistic coupling to variables in other modalities, such as action planning, motion control, or language. It is these explicit, temporarily committed representations that give us the feeling of being in direct contact with the visual reality out there.

SURELY, OUR MODEL STILL LEAVES OPEN A NUMBER OF IMPORTANT PROBLEMS FOR FUTURE WORK. AS IS, THE MODEL IN Variant only to translation and needs to be generalized to changing scale and orientation (which will require dynamic relinking of feature connections, see Sato, Jitsev, & von der Malsburg, 2008) as well as other image transformations such as changing illumination and perspective deformation. As discussed before, the system as proposed here assumes direct dynamic links from all positions in the Input Layer to all positions of the Input Assembly. This would require unrealistic fiber convergence numbers. Following the proposal of Olshausen et al. (1993), this problem can be solved with the help of a switchyard of connections over several intermediate layers. An optimized system of this kind (Wolfrum & von der Malsburg, 2007b) has reasonable complexity (of $O(n \log n)$ with a small prefactor) in terms of numbers of connections and intermediate units. A further problem yet to be solved is the generation of the gallery domain with the help of autonomous learning mechanisms and the generation of the very specific connectivity patterns of the link control units by ontogenetic mechanisms. A possibly unrealistic aspect of our model in its present form is the maintenance of the same low-level feature types (Gabor wavelets) through large parts of the system, a hierarchy of feature types of growing complexity being more likely to be realized in the visual system. Incorporation of this will most naturally be realized in the context of a more comprehensive model encompassing also ontogenesis and learning mechanisms. Another unrealistic aspect of the model is the flat structure of the gallery domain. As has been convincingly shown by Biederman (1987), many objects are recognized as ordered arrays of simpler subshapes. Again, realization of such structure in the gallery domain will require potent learning mechanisms.

Object recognition is only one of a great multitude of functions of the visual system and of the brain. We feel, however, that the principle of correspondence finding by active information routing is of general importance as a paradigm of brain function.

Appendix A

Parameters

The size of the network is determined by the following parameters. The input grid consists of $P = 400$ columns, all face graphs (Assembly and Gallery Layers) contain $Q = 48$ columns. Consequently, there is a total of $400 \times 48 = 19200$ control units. For representation of visual information in the Input and Assembly Layers, we use $K = 40$ Gabor wavelets. The number $M$ of Gallery faces depends on the size of the database on which the system is tested. In the case of the FERET database, this is $M_{FERET} = 1196$, for the AR database we have $M_{AR} = 100$.

For the simulation results of the Simulations section we used the following settings and parameters. We chose a time constant of $\tau = 0.2 \text{ ms}$. As mentioned in the Simulations section, the specific choice of $\tau$ does not influence the behavior of the system as long as the ratio $\frac{L}{T}$ between recognition time and time constant is fixed. The length of the $\nu$ cycle we chose is $T = 400 \times 80 \text{ ms}$. The radius for topological interaction among control units was $B = 0.05 \times \text{image size}$. Maximum strength of this topological interaction was $c_{top,c} = 3.5$. Cooperation strength between neighboring gallery units was $c_{top,g} = 0.1$. 
Proof of self-normalization properties

Since the ratio \( \frac{C}{C_1} \) is very large, i.e., the time constant is much shorter than the overall simulation time, the unit activities are close to the adiabatic state, i.e., \( \frac{d}{dt}x_i \approx 0 \). We therefore analyze the steady state of the unit activities to derive the self-normalization properties.

For the case of \( \nu = 0 \) (linear representation), the unit dynamics are given by Equation 2:

\[
\tau \frac{d}{dt} x_i = I_i - x_i \sum_{j=1}^{K} I_j x_j. \tag{A1}
\]

The steady state of this is

\[
x_i = \frac{I_i}{\sum_j I_j x_j}. \tag{A2}
\]

Multiplying both sides with \( x_i \) yields

\[
x_i^2 = \frac{I_i x_i}{\sum_j I_j x_j}, \tag{A3}
\]

and the sum of this term over all \( i \) is

\[
\sum_i x_i^2 = 1, \tag{A4}
\]

i.e., for \( \nu = 0 \) the column activity is normalized to a 2-norm of 1.

In the case of \( \nu = 1 \) (WTA behavior), the unit dynamics are

\[
\tau \frac{d}{dt} x_i = x_i I_i - x_i \sum_{j=1}^{K} I_j x_j, \tag{A5}
\]

so in steady state we have

\[
\tau \sum_i \frac{d}{dt} x_i = \sum_i x_i I_i - \sum_i x_i \sum_j x_j I_j = \left(1-\sum_i x_i \right) \sum_i x_i I_i = 0. \tag{A6}
\]

If there is any activity in the column, i.e., not all unit activities are zero simultaneously, this requires

\[
\sum_i x_i = 1, \tag{A7}
\]

which means that for \( \nu = 1 \) the column activity self-normalizes to a 1-norm of 1. Consequently, the interaction term \( \sum_j I_j x_j \) is the average activity-weighted input to the column. This means that only those unit activities grow whose input is higher than this weighted mean input to the column, otherwise they decay. This lets the weighted input average grow, because the bias shifts toward strong inputs. Eventually, all unit activities decrease to 0 except for the unit with the strongest input, whose activity approaches 1. For the final steady state we can show this by setting the time derivative for a single unit to 0:

\[
x_i \left(I_i - \sum_{j=1}^{K} I_j x_j \right) = 0. \tag{A8}
\]

Here we see that for any \( i \), we either have \( x_i = 0 \), or \( I_i = \sum_j I_j x_j \), which can only be true for \( x_i = 1 \) and all other \( x_j = 0 \) (except for the degenerate case of two or more of the \( I_i \) being exactly identical, which would presumably be solved in the brain by spontaneous symmetry breaking).

Gabor transform

The model described in this paper uses a set of Gabor wavelets that appropriately sample orientation (over 8 orientations) and spatial frequency (over 5 scales) space. If \( V \) is an image with \( V(\vec{z}) \) denoting the gray value of a pixel at the geometric position \( \vec{z} = (x_y) \), the filter responses \( R_i(\vec{z}) \) are given by

\[
R_i(\vec{z}) = \int V(\vec{z}') \psi_i(\vec{z} - \vec{z}') d\vec{z}', \tag{A9}
\]

\[
\psi_i(\vec{z}) = \frac{k_i^2}{\sigma^2} \exp\left(-\frac{k_i^2 \vec{z}^2}{2\sigma^2}\right) \left[\exp\left(i k_i^y \vec{z}^y\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right], \tag{A10}
\]

where the wave vector is parameterized as

\[
\vec{k}_i = \begin{pmatrix} k_i^x \\ k_i^y \end{pmatrix} = \begin{pmatrix} k_p \cos \varphi_\mu \\ k_p \sin \varphi_\mu \end{pmatrix}, \quad k_p = 2\left(\frac{\sigma^2}{2}\right)^{\frac{1}{2}}, \quad \varphi_\mu = \frac{\pi}{8}, \tag{A11}
\]

with orientation parameter \( \mu = 1, \ldots, 8 \) and scale parameter \( \rho = 1, \ldots, 5 \). That is, \( R_i(\vec{z}) \) is a vector of Gabor filter responses in which each entry corresponds to one of the 40 combinations of \( \rho \) and \( \mu \). As feature values we use the magnitude

\[
J_i^\rho = |R_i(\vec{z}_p)|, \tag{A12}
\]

thus ignoring Gabor phase.
Acknowledgments

We thank Urs Bergmann for his help in programming, Alexander Heinrichs for helping to preprocess the database images, Cornelius Weber and Junmei Zhu for useful comments on the manuscript, and two anonymous reviewers whose criticism has helped us to clarify a number of points. This work was supported by the European Union through Project FP6-2005-015803 (“Daisy”), by the Gatsby Charitable Foundation, and by the Hertie Foundation.

Commercial relationships: none.

Corresponding author: Philipp Wolfrum
Email: wolfrum@fias.uni-frankfurt.de
Address: FIAS, Ruth-Moufang-Str. 1, 60433 Frankfurt, Germany.

Footnotes

1 Note that the concept of a canonical microcircuit is not necessarily tied to the concept of anatomically disjunct columns as promoted in Mountcastle (1997).
2 In principle, the competition parameter \( \nu \) could be set to a constant value of \( \nu = 1 \). However, slowly increasing competition within the columns of a network has in earlier systems proven to efficiently avoid local optima (Lücke et al., 2008). This is related to the slow change of the temperature parameter in simulated annealing-like systems (Kirkpatrick, Gelatt, & Vecchi, 1983), which serves the same purpose. In Kömer, Gewaltig, Kömer, Richter, and Rodemann (1999) the thalamic complex of intralaminar nuclei is discussed as a possible source of a fast and global modulatory signal to the cortex.
3 For brevity of notation, we will sometimes just use the name of a certain unit type (like \( C \) for control units) to denote the output of that unit. We will always point this out when we do so.

References


