Modelling supply networks and business cycles as unstable transport phenomena

Dirk Helbing
Institute for Economics and Traffic, Dresden University of Technology, Andreas-Schubert Straße 23, 01062 Dresden, Germany
E-mail: helbing@trafficforum.org

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Received 12 May 2003
Published 14 July 2003

Abstract. Physical concepts developed to describe instabilities in traffic flows can be generalized in a way that allows one to understand the well-known instability of supply chains (the so-called ‘bull-whip effect’). That is, small variations in the consumption rate can cause large variations in the production rate of companies generating the requested product. Interestingly, the resulting oscillations have characteristic frequencies which are considerably lower than the variations in the consumption rate. This suggests that instabilities of supply chains may be the reason for the existence of business cycles. At the same time, we establish some links to queueing theory and between micro- and macroeconomics.
1. Introduction

Concepts from statistical physics and non-linear dynamics have been very successful in discovering and explaining dynamical phenomena in traffic flows [1]. Many of these phenomena are based on mechanisms such as delayed adaptation to changing conditions and the competition for limited resources, which are relevant for other systems as well. This includes pattern formation such as segregation in driven granular media [2] and lane formation in colloid physics [3] or biological physics (pedestrians, ants) [4]. Other examples are clogging phenomena at bottlenecks in freeway traffic [5], panicking pedestrian crowds [6], or granular media [7]. In the following study, we will focus on the phenomenon of stop-and-go traffic [8] and its analogies.

Recently, economists and traffic scientists have wondered whether traffic dynamics also has implications for the stability and management of supply chains [9, 10] or for the dynamics of business cycles [11]. To explain business cycles, many theoretical concepts have been suggested over the decades, such as the Schumpeter clock [12]. These are usually based on macroeconomic variables such as investment, income, consumption, public expenditure or the employment rate, and their interactions. In contrast, Witt et al [11] have recently suggested interpreting business cycles as self-organization phenomena due to a linear instability of production dynamics related to stop-and-go waves in traffic or driven many-particle systems. In order to illustrate their idea,
they have transferred a continuous macroscopic traffic model and reinterpreted the single terms and variables.

We believe that this is a very promising approach to understand business cycles, but instead of simply transferring macroscopic traffic models we suggest derivation of the equations for business cycles from first principles, which means deriving the dynamics on the macroscopic level from microscopic interactions in production systems. This would also make some contribution to the goal of understanding macroeconomics based on microeconomics (or, in a wider sense, based on the ‘elementary interactions’ of individuals; here, production managers).

In order to make some progress in this direction, we will generalize [13] an idea suggested by Daganzo to describe the dynamics of supply chains [9]. Like the work by Armbruster et al [10], his approach is related to traffic models as well, but he focuses on models in discrete space in order to reflect the discreteness of successive production steps. In order to describe the non-linear dynamics of production processes or interrelated economic sectors, we will have to generalize these ideas to complex supply networks. In section 2, we will first discuss ‘macroscopic’ business cycles in a sectorally structured economy and compare them with stop-and-go traffic. Afterwards, in section 3, we will develop a more fine-grained, ‘microscopic’ description of the management of dynamically interacting production units and relate it to classical queueing theory (which mainly focuses on stochastic fluctuations of production processes in a stationary state). Later on, in section 4, we will construct a mathematical relation between the microscopic and the macroscopic level of description, while some further research directions and other potential applications are indicated in the outlook of section 6.

2. Modelling ‘macroscopic’ supply networks

2.1. Economic production sectors

We will first investigate a simplified economic system with $U$ production sectors $B$ generating certain kinds of products $I \in \{1, \ldots, P\}$. The market for products $I$ as a function of time $t$ will be represented by the stock level (‘inventory’) $N_I(t)$. Let us assume that, in each production cycle, the production sector $B$ generates $p_B^I$ products of kind $I$ and requires $c_B^J$ products of kind $J$ (‘educts’) for production. If $Q_B(t)$ is the number of production cycles per unit time, i.e. a measure of the production rate, productivity or ‘throughput’ of sector $B$, the quantity of products generated per unit time is $p_B^I Q_B(t)$, while the quantity of educts consumed per unit time is $c_B^J Q_B(t)$ (see figure 1). The temporal change of the quantity of products $I$ in the market is, therefore, given by the conservation equation

$$\frac{dN_I}{dt} = \sum_B (p_B^I - c_B^I) Q_B(t).$$

(1)

The production rate $Q_B(t)$ will be specified later on in section 3. For the time being, we will assume that $Q_B(t)$ agrees with the actual feeding (‘arrival’) rate $\lambda_B$ discussed in the next subsection:

$$Q_B(t) = \lambda_B(t).$$

(2)
2.2. The feeding rates

In the absence of capacity constraints, we would have the relation \( \lambda_B = \rho^0_B \mu_B \) for the time-dependent feeding rate \( \lambda_B \), where \( \mu_B \) is the processing rate of sector \( B \), i.e. the potential production rate in the case of no inefficiencies. \( \rho^0_B \) reflects the desired utilization. A value of \( \rho^0_B = 0.7 \) is reasonable, as it guarantees a relatively high production rate at moderate and reliable waiting times for finishing the products (see section 3.1). However, the actual utilization \( \rho_B = \lambda_B / \mu_B \) does not necessarily agree with the desired one, \( \rho^0_B \). In order to illustrate this, let us assume that \( T_B^J \) is the average time needed to deliver to production sector \( B \) products of kind \( J \), which are required for production (they will be called ‘educts’ in the following). If \( N_J \) is the quantity of actually available products of kind \( J \), the resulting maximum quantity of products that can be delivered per unit time is \( N_J / T_B^J \). As \( c_B^J \) educts are required per production cycle, one cannot have a higher production rate than \( N_J / (T_B^J c_B^J) \) per unit time, which defines the maximum feeding rate \( \lambda_B^J \) of educt \( J \) into production unit \( B \):

\[
\lambda_B^J = N_J / (T_B^J c_B^J). \tag{3}
\]

Moreover, as the products can only be finished if all the required educts are available in a production cycle, the actual feeding rate \( \lambda_B \) is given by the minimum of these values and of the desired feeding rate \( \rho^0_B \mu_B \) (see figure 1):

\[
\lambda_B(t) = \min(\rho^0_B \mu_B, \{\lambda_B^J\}). \tag{4}
\]
If $c_J^B = 0$, we set $\lambda_J^B \to \infty$, so that the corresponding term does not have any impact on the value of the minimum function (4).

Note that production has some analogy to chemical reactions, where one also requires a certain quantity of educts to produce other (chemical) products. For chemical reactions in three-dimensional space, however, the reaction rate is given by multiplication of (a power of) the chemical concentrations of the educts, i.e. the above minimum function is replaced by an algebraic product. This difference is comparable with the probabilistic AND and the fuzzy AND in fuzzy logic, which are applied when several conditions are to be met at the same time. The probabilistic AND is based on a multiplication of the logical values, while the fuzzy AND corresponds to their minimum.

2.3. Adaptation of production speeds and transport capacities

One important aspect is the adaptation of production to the time-dependent inventories $N_J(t)$. The adaptation of the desired utilization $\rho^0_B$ is delayed (see section 3), and a change of the processing rate $\mu_B$, which requires an adaptation of production sectors or capacities, takes a considerable time as well. We will assume

$$\frac{d(\rho^0_B \mu_B)}{dt} = \frac{1}{\tau_B} [W_B(1/Z_B, \ldots) - \rho^0_B \mu_B]$$

with a typical adaptation time $\tau_B$. $W_B(1/Z_B, \ldots)$ reflects the production rate of sector B at steady state as a function of the inventories $N_J$. For the time being, we will assume

$$\frac{1}{Z_B(t)} = \sum_J p_J^B X_J^B N_J(t),$$

i.e. the production speed is adapted only to the inventories $N_J$ of the products $J$ generated by sector $B$, for which $p_J^B > 0$. $X_J^B$ is an additional and constant prefactor. This formula is defined in a way which always gives positive values $Z_B(t)$, if not all coefficients $p_J^B$ are zero and the $N_J(t)$ stay positive (see below). Normally, the production speed $W_B(1/Z_B, \ldots)$ will increase with decreasing inventories $N_J$, but it saturates due to financial or technological limitations and inefficiencies. In this study, we will assume

$$W_B(1/Z_B, \ldots) = \max\left(A_B \frac{1 + B_B Z_B}{1 + B_B Z_B + D_B(Z_B)^2}, 0\right),$$

where $Z_B$ could be called ‘scaled inventories’ and $A_B$, $B_B$ and $D_B$ are suitable parameters (see figure 2).

Furthermore, we will assume here that the transport capacities are adapted in parallel with the production speed and represent the proportionality constants by $V_J^B$. This implies

$$\lambda_J^B = \rho^0_B \mu_B V_J^B N_J/c_J^B,$$

corresponding to the specification $1/T_J^B = \rho^0_B \mu_B V_J^B$, which results in the following formula for the feeding rates:

$$\lambda_B(t) = \rho^0_B \mu_B \min\left(1, \left\{\frac{V_J^B N_J}{c_J^B}\right\}\right).$$

Herein, the minimum extends over all indices $J$. 

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
2.4. Bull-whip effect and stop-and-go traffic

For a linear supply chain with $c_I^t = \delta_{B, I+1}$ and $p_I^t = \delta_{B, I}$ (where $\delta_{K, L} = 1$ if $K = L$, and $\delta_{K, L} = 0$ otherwise) we obtain the particular set of equations

\[
\frac{dN_I}{dt} = Q_I(t) - Q_{I+1}(t) = \lambda_I(t) - \lambda_{I+1}(t).
\]

(10)

In the capacity-constrained case (characterized by small values of $V^I_B$), this leads to

\[
\frac{dN_I}{dt} = V_I(t)N_{I-1}(t) - V_{I+1}(t)N_I(t)
\]

(11)

with

\[
V_I(t) = \rho_0^I(t)\mu_I V_{I-1}^I
\]

(12)

and

\[
\frac{d(\rho_0^I\mu_I)}{dt} = \frac{1}{\tau_I}[W_I(X_I^I/N_I) - \rho_0^I\mu_I]
\]

(13)

or

\[
\frac{dV_I}{dt} = \frac{1}{\tau_I}[V_{I-1}^I W_I(X_I^I/N_I) - V_I(t)].
\]

(14)

Interestingly, equations (11) and (14) basically agree with the macroscopic traffic flow model by Hilliges and Weidlich [14], where (11) is analogous to the equation for the vehicle density and (14) corresponds to the equation for the average vehicle speed $V_I$ in street segment $I$. These equations are linearly unstable with respect to perturbations of $N_I(t)$, if $\tau_I$ exceeds a certain threshold (see figure 3), which depends on the maximum slope $W'(X_I^I/N_I)$ of $W(X_I^I/N_I)$ [14]. As drivers (over-)react with a time delay to a changing traffic situation in front, stop-and-go traffic can emerge. The frequently observed instability of supply chains, called the ‘bull-whip effect’, occurs for similar reasons (e.g. in the ‘beer distribution game’ [15, 16]) (see figure 5).

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Figure 3. Minimum and maximum inventories as a function of the adaptation time $\tau_B$. The other model parameters are specified as in figure 5, but the perturbation amplitude has been chosen five times smaller. The difference between both curves is the amplitude of the bull-whip effect, i.e. the maximum time-dependent variation in the inventories. Note that there is a critical adaptation time, below which perturbations are not amplified. In this case, the investigated linear supply chain is stable.

The mechanism behind this instability is the delay $\tau_I$ in the adaptation of the production speeds and transport capacities, which implies over- or underproduction. The repeatedly or periodically resulting high inventories are due to temporary bottlenecks in the supply chain and could be avoided by appropriate control functions $W_B = W_B([N_I(t)], \{dN_I/dt(t)\}, \ldots)$ [17, 18]. Note that, instead of unstable production with relatively high inventories and low production speeds, one may reach the same average, but stable throughput at low inventories and high production speeds (see figure 4). However, this is normally related to higher energy and maintenance costs, so that production tends to operate in the linearly unstable regime.

If the transport capacities are not a limiting factor (i.e. the parameters $V_B^I$ are large), instead of (11) we have the set of equations

$$\frac{dN_I}{dt} = \rho_I^0(t)\mu_I - \rho_{I+1}^0(t)\mu_{I+1}. \quad (15)$$

Together with (14), this basically corresponds to a particular microscopic traffic model, the so-called optimal velocity model [19], if we restrict our comparison to the linear regime around the steady state and identify $\rho_I^0\mu_I$ with the actual velocity of vehicle $I$, but $X_I^I/N_I$ with the inverse distance to the next car ahead (apart from proportionality constants). It is known [17, 19] that this model is linearly unstable if

$$V_I^{I-1}W_I'(X_I^I/N_I) > \frac{1}{2\tau_I}, \quad (16)$$

i.e. if the adaptation time $\tau_I$ exceeds a certain threshold which depends on the slope $W_I'(X_I^I/N_I)$ of $W_I(X_I^I/N_I)$. 

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Figure 4. Steady-state throughput $Q_B$ as a function of the scaled inventories $Z_B$ for the parameter $B_B = 0.2$, arbitrary $A_B$, and various values of $D_B$ in the capacity-restricted case of low transport rates $V'_B$. For large enough values of $D_B$, the curve has a maximum at finite inventories. In this case, high steady-state flows $Q_B$ appear twice: for low (and linearly stable) inventories and higher (but potentially unstable) inventories (see also figure 6).

2.5. Business cycles

The most interesting point is the reaction of the system to a perturbation in the throughputs $Q_B(t)$, e.g. a periodic perturbation of the so-called consumption rate $Q_{U+1}(t)$ (see sections 4.3 and 4.4). The resulting oscillations in the inventories can be much slower and are usually synchronized among different production sectors (see figures 5, 11 and 12(a)). Therefore, they may explain business cycles as a self-organized phenomenon with slow dynamics on the time scale of several years.

The reduction in the resulting oscillation frequency compared with the perturbation frequency is also known from stop-and-go traffic. It has been explained by the non-linearity in the model equations. However, there is a significant difference between the dynamics of traffic flows and supply chains: while stop-and-go waves have a characteristic amplitude independently of the average vehicle density, the amplitude of oscillations in the inventories changes continuously in the capacity-constrained case (see figure 6). In other words, the phase transition from stable to unstable traffic flow is hysteretic (i.e. of first order) [20, 21], while the phase transition from stable to unstable supply chains in the capacity-constrained case appears to be continuous (i.e. of second order). This seems to be a particular property of the Hilliges–Weidlich model, while the optimal-velocity model mentioned in section 2.4 is known to display a hysteretic transition [19]. Therefore, a hysteretic transition is found for supply chains which are not constrained by their transport capacities. The transition between these two different regimes would be interesting to investigate.

Note that, in order to find the emergence of slow oscillations, i.e. of business cycles, we do not need to have a linear supply chain. Supply networks can display similar features (see section 5). The only requirement is that no stationary state exists or the stationary state is linearly unstable with respect to perturbations. This is being intensively studied for particular network types in a recent scientific collaboration [18].
Figure 5. Variations of the inventories $N_I(t)$ of $U = P = 10$ economic sectors $I$, triggered by a constant consumption rate with a slight periodic perturbation, namely $V_{11}(t) = V_{11}^{10}W_I(1)[1 + 0.1 \sin(0.1t)]$. One can clearly see the emergence of slow and synchronized oscillations of the inventories $N_I(t)$, which are triggered by small and fast perturbations in the consumption rate $V_{11}(t)$. The unit time is one day and the model parameters are $A_B = A_I = 100/V$, $V_I^{1-1} = V = 10^{-4}$, and $\tau_B = \tau_I = 90$. The initial and boundary conditions are $N_I(0) = N_0(t) = 20 = X_I^0$, $V_I(0) = V W_I(X_I^0/N_I(0)) = V W_I(1)$, and $V_{11}(t)$ (see above).

Figure 6. Minimum and maximum inventories as a function of the scaled steady-state inventory $Z_I(0)$. The other model parameters are specified as in figure 5, but the perturbation amplitude has been chosen five times smaller (of the order of 2%). The difference between both curves reflects the amplitude of the maximum time-dependent variation in the inventories. Note that, in the capacity-constrained Hilliges–Weidlich case investigated here, the change of the amplitude (and the associated kind of non-equilibrium phase transition) is continuous. The supply chain is stable with respect to perturbations, where both curves agree. This is the case at small and large steady-state inventories.
Summarizing our present insights, preconditions for the emergence of business cycles are

- large values of the adaptation times $\tau_B$,
- significant changes of the production speeds $W_B$ with changing inventories,
- the presence of perturbations, if a steady system state exists,
- no consideration of forecasts or poor forecasts of the future time development of the inventories $N_I(t)$,
- a highly non-linear interaction among the production sectors.

We should finally emphasize that the above considerations can be applied to economic systems with any number of sectors and any network structure. We do not even need to assume a sectorally structured economy, as the same kinds of equations apply on the ‘microscopic’ level of production networks, only with a significantly higher number of equations (see section 4.1). The special cases discussed above have been chosen only for illustrative reasons.

3. ‘Microscopic’ model of production processes

In this section, we will formulate a generalized model for the dynamical interaction of production processes. Compared with the description of interacting economic sectors in the previous section, this approach may be called microscopic. Note that the concept developed in this section is required for two reasons: first in order to be able to describe real production processes, which involves buffers and other variables and calls for a more complex model; second in order to allow the derivation of the macroeconomic dynamics from microeconomic assumptions regarding the production management of single companies (see section 4). Readers not interested in these aspects may skip this section and continue with section 5.

3.1. Production units in terms of queueing theoretical quantities

We will now investigate a system with $u$ production units (machines or factories) $b \in \{1, 2, \ldots, u\}$ producing $p$ products $i, j \in \{1, 2, \ldots, p\}$. The respective production process is characterized by parameters $c^i_b$ and $p^j_b$: in each production cycle, production unit $b$ requires $c^j_b$ products (‘educts’) $j \in \{1, \ldots, p\}$ and produces $p^i_b$ products $i \in \{1, \ldots, p\}$. The number of production cycles of production unit $b$ per unit time is a measure of the throughput and will be represented by $Q_b(t)$. For production in the steady state, we can often approximate this quantity in terms of variables from queueing theory [22]. For this, let $\lambda_b$ be the feeding (‘arrival’) rate, $C_b$ the number of parallel production channels, $\mu_b$ the overall processing (‘departure’) rate (i.e. $C_b$ times the processing rate of a single production channel),

$$\rho_b = \lambda_b / \mu_b$$

(17)

the utilization and $S_b$ the storage capacity of production unit $b$. Then the following relationships apply (see figure 7):

$$Q_b(t) = (1 - p^\mu_b)\lambda_b(t) = (1 - p^\lambda_b)\mu_b(t).$$

(18)

The functions $p^\lambda_b$ and $p^\mu_b$ reflect a reduction in the efficiency of the feeding (arrival) and the processing (departure) in production unit $b$. They depend on $\rho_b$, $C_b$, $S_b$ and possibly other
Figure 7. Schematic illustration of a production unit \( b \) as a queueing system with a limited storage capacity \( S_b \) and \( C_b \) parallel production channels. The arrival rate \( \lambda_b \), the departure rate \( \mu_b \) and the effects of inefficiencies are indicated.

quantities as well. In some cases, \( p_b^\lambda \) is the probability of rejecting arrivals when the storage capacity \( S_b \) is reached, i.e.

\[
p_b^\lambda(\rho_b, C_b, S_b) = P_b(S_b), \tag{19}
\]

where \( P_b(l) \) is the probability of having a queue of length \( l_b \) in production unit \( b \). On the other hand, the average number of active production channels is

\[
\sum_{l_b=0}^{C_b-1} l_b P_b(l_b) + \sum_{l_b=C_b}^{S_b} C_b P_b(l_b). \tag{20}
\]

Consequently, the relative reduction \( p_b^\mu \) in the processing rate due to \( (C_b - l_b) \) empty channels is

\[
p_b^\mu(\rho_b, C_b, S_b) = \sum_{l_b=0}^{C_b-1} \frac{C_b - l_b}{C_b} P_b(l_b). \tag{21}
\]

In more complex production systems, the factors \( p_b^\lambda \) and \( p_b^\mu \) would be generalized measures of inefficiencies in the production process, which would be functions of all relevant process parameters.

According to Little’s law for queueing systems, \( L_b = \lambda_b T_b \), where \( \lambda_b = (1 - p_b^\lambda)\lambda_b \) is the average arrival rate, we can express the throughput also in terms of the average queue length \( L_b(\rho_b, C_b, S_b) \) and the average waiting time \( T_b(\rho_b, C_b, S_b) \):

\[
Q_b(t) = \frac{L_b(\rho_b, C_b, S_b)}{T_b(\rho_b, C_b, S_b)}. \tag{22}
\]

For example, for a system with infinite storage capacity \( S_b \rightarrow \infty \), we have

\[
Q_b = \min(\lambda_b, \mu_b) = \mu_b \min(\rho_b, 1). \tag{23}
\]

For an \( M/M/1:(S_b/FIFO) \) process (one channel with first-in–first-out serving, storage capacity \( S_b \), Poisson-distributed arrival times and exponentially distributed service intervals), one finds for \( \lambda_b \leq \mu_b \)

\[
Q_b = \lambda_b \frac{1 - \rho_b S_b}{1 - \rho_b^{S_b+1}} \left( \frac{\rho_b}{1 - \rho_b S_b+1} \right) \frac{S_b}{S_b + 1}. \tag{24}
\]

Note that both the expected value and the standard deviation of the queue length and the waiting time diverge for \( \rho_b \rightarrow 1 \). Therefore, efficient production is normally related to \( \rho_b \leq 0.7 \) [22].

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
3.2. The feeding rates of production units

In the absence of capacity constraints, we just have the relation $\lambda_b = \rho_0^b \mu_b$ for the feeding rate, where the actual utilization $\rho_b$ agrees with the desired utilization $\rho_0^b$, e.g. $\rho_b = 0.7$. However, a lack of required educts may lead to a reduction of $\lambda_b$. In that case, the feeding rate is limited by the minimum arrival rate $A^j_b$ of required educts $j$, divided by the quantity $c^j_b$ of educts needed for one production cycle. We will assume that the maximum arrival rates $A^j_b$ are given by $\rho_b^0 \mu_b V^j_b I^j_b$, where $I^j_b$ denotes the quantity of educts stored in the input buffer, $\rho_b^0 \mu_b V^j_b$ is the maximum transport rate due to capacity constraints $V^j_b$ for getting educt $j$ from the input buffer into the production unit $b$, and the prefactor $\rho_b^0 \mu_b$ suggests that the transport capacity is adapted to the production speed (generalizations are possible). Therefore, if $\lambda^j_b = \rho_0^b \mu_b V^j_b I^j_b / c^j_b$ denotes the maximum feeding rate for educt $j$ into production unit $b$, the actually resulting feeding rate is

$$\lambda_b(t) = \min(\rho_0^b \mu_b, \{\lambda^j_b\}),$$

where the minimum extends over all indices $j$. The product flows are illustrated in figure 8.

When we assume that the flows $S^j_{ab}$ of products $j$ to $b$ from the delivering production units $a$ can be fed into the production process in parallel (instead of having to go through the input
buffer first), we have the generalized relationship

\[
\lambda_b^j(t) = \left( \rho_b^0 \mu_b V_b^j I_b^j(t) + \varepsilon \sum_a S_{ab}^j(t) \right) / c_b^j
\]

(26)

with \( \varepsilon = 1 \). The previously discussed case (requiring delivery through the input buffer) corresponds to \( \varepsilon = 0 \). In any case, the fraction of the supply \( \sum_a S_{ab}^j \), which is not needed to satisfy the production requirements \( c_b^j Q_b \), is delivered to the input buffer.

### 3.3. Input and output buffers

Let us assume each production unit \( b \) has input buffers for required products \( i \) (‘educts’) and output buffers \( a \) (‘warehouse’) for the products. We will assume the input buffers are filled with \( I_b^i(t) \) educts \( i \in \{1, \ldots, p\} \) and the output buffers with \( O_b^j(t) \) products waiting to be delivered. If \( S_{ab}^j(t) \) denotes the delivery flow (‘supply’) of products \( i \) from production unit \( a \) to \( b \), the change of an input buffer stock with time is given by the conservation equation

\[
\frac{dI_b^i}{dt} = \sum_a S_{ab}^j(t) - c_b^j Q_b(t),
\]

(27)

as \( \sum_a S_{ab}^j(t) \) is the quantity of products \( i \) delivered from various sources (production units) \( a \), and \( c_b^j Q_b(t) \) is the quantity of educts \( i \) used up for production per unit time (see figure 8). Analogously, the dynamics of an output buffer stock is determined by the equation

\[
\frac{dO_b^j}{dt} = p_b^j Q_b(t) - \sum_c S_{bc}^j(t),
\]

(28)

as \( p_b^j Q_b(t) \) is the quantity of newly generated products \( j \), and \( \sum_c S_{bc}^j(t) \) are deliveries to other production units \( c \) (see figure 8). The following specifications in this paper ensure the non-negative conditions \( I_b^i(t) \geq 0 \) and \( O_b^j(t) \geq 0 \) (see also sections 3.5 and 4.1). We will, for example, assume that the delivery flows \( S_{ab}^j \) are basically given by the order flows (‘demands’) \( D_{ab}^j \), but limited by the maximum transport rates \( V_{ab} \rho_b^0 \mu_b \) for delivering the available products from output buffer \( O_b^j \) to production unit \( b \):

\[
S_{ab}^j(t) = \min[D_{ab}^j(t), O_b^j(t) V_{ab} \rho_b^0 \mu_b].
\]

(29)

If required, suitable specifications can also guarantee \( I_b^i(t) \leq I_b^{i,\max} \) and \( O_b^j(t) \leq O_b^{j,\max} \) with maximum input and output buffer stocks \( I_b^{i,\max} \) and \( O_b^{j,\max} \). We may, for example, multiply \( S_{ab}^j \) by \( [1 - (I_b^i/I_b^{i,\max})^\kappa] \), which stops the delivery flow of product \( i \) when the respective input buffer of production unit \( b \) is full. Analogously, we may multiply \( \rho_b^0 \mu_b \) by \( \prod_j [1 - (O_b^j/O_b^{j,\max})^\kappa] \), which stops the production when one of the output buffers of \( b \) is full. High values of \( \kappa \) produce a hard cutoff, while small values of \( \kappa \geq 1 \) can describe cases where the production efficiency goes down even before the buffer size is fully used up.

### 3.4. Adaptation of production speeds and transport capacities

One important aspect is the adaptation of production to changing demand. On the one hand, a change of the processing rate \( \mu_b \) is expensive and time consuming. On the other hand, the

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
adaptation of the desired utilization $\rho_0^b$ is delayed by the average queueing time $T_b$ and other factors. We will assume
\[
\frac{d(\rho_0^b \mu_b)}{dt} = \frac{1}{\tau_b} [W_b(1/Z_b, \ldots) - \rho_0^b \mu_b]
\] (30)
with a typical adaptation time $\tau_b$ significantly greater than the adaptation times of the other variables. (That is why we do not consider possible time delays in the other mathematical relations, here.) $W_b(1/Z_b, \ldots)$ is some control function reflecting the strategy by the production manager in adapting the utilization and/or processing rate to the output buffer stocks $O_j^a$ or other variables. In the following, we will assume
\[
\frac{1}{Z_b} = \sum_j p_j^b X_j^b/N_j^b,
\] (31)
i.e. the management strategy is only sensitive to the perceived stock levels $N_j^b$ of the products $j$ generated by unit $b$, for which $p_j^b > 0$. $X_j^b$ are proportionality constants. The perceived stock levels will be specified by
\[
N_j^b = O_j^b + \delta \sum_{a \neq b} O_j^a.
\] (32)
For $\delta = 0$, the management takes into account only its own output buffer stock $O_j^b$, while for $\delta = 1$, it also tracks the output buffer stocks $O_j^a$ of competing production units $a$. Normally, the control function $W_b$ will increase with decreasing perceived stock levels $N_j^b$, but the function $W_b(1/Z_b, \ldots)$ saturates due to financial, spatial, or technological limitations and inefficiencies in the processing of high order flows. Again, we will use a function of the form
\[
W_b(1/Z_b, \ldots) = \max \left( \frac{1}{Z_b} \frac{1 + B_b Z_b}{1 + B_b Z_b + D_b(Z_b)^2}, 0 \right)
\] (33)
with suitable parameters $A_b, B_b$ and $D_b$. $A_b$ corresponds to the maximum production speed.

3.5. Order flows, delivery networks and price dynamics

The flow of orders is basically given by the flow $c_i^b Q_b$ of educts required for the production by unit $b$. Deviations can be reflected by a correction function $U_b(I_t^b, I_{t0}^b)$, where $I_{t0}^b$ denotes the desired input buffer stock. If the proportions $q_{ab}^i$ with $\sum_a q_{ab}^i = 1$ reflect orders from different producers $a$, we have
\[
D_{ab}^i = q_{ab}^i c_i^b Q_b U_b(I_t^b, I_{t0}^b)
\] (34)
and, therefore,
\[
S_{ab}^i(t) = \min[q_{ab}^i c_i^b Q_b U_b(I_t^b, I_{t0}^b), O_a^i V_a^i \rho_0^b \mu_b]
\] (35)
(see equation (29)). When the first term is the smaller one, this implies with (27)
\[
\frac{dI_t^b}{dt} = \sum_a q_{ab}^i c_i^b Q_b U_b(I_t^b, I_{t0}^b) - c_i^b Q_b
\]
\[
= [U_b(I_t^b, I_{t0}^b) - 1] c_i^b Q_b(t).
\] (36)
The function \( U_b(I^i_b, I^0_b) \) should be chosen greater than one if \( I^i_b \) is smaller than the desired input buffer stock \( I^0_b \), smaller than one for \( I^i_b > I^0_b \), and otherwise one, e.g.

\[
U_b(I^i_b, I^0_b) = I^0_b / I^i_b.
\] (37)

Without this correction function, the input buffer tends to be emptied in the course of time.

Together with \( c^i_b \) and \( p^{i'}_{ab} \), the fractions \( q^i_{ab} \) characterize the delivery or supply network. One can imagine various scenarios. For example, the fractions could be assumed to be constant or modelled by an evolutionary selection equation with a selection rate \( \nu \) [23]:

\[
\frac{d q^i_{ab}}{dt} = \nu (F^i_{ab}(t) - \sum_{a'} F^{i'}_{a'b}(t) q^{i'}_{a'b}(t)) q^i_{ab}(t).
\] (38)

The specification of the fitness \( F^i_{ab} \) depends on the relevant parameters. For example, it could be treated as constant. However, in some cases, it makes sense to relate the fitness \( F^i_{ab} \) to the inverse of the real or virtual costs (‘prices’) \( p^{i}_{ab} \) of product \( i \), when delivered from production unit \( a \) to \( b \):

\[
F^i_{ab}(t) = 1 / p^{i}_{ab}(t).
\] (39)

Assuming a law of supply and demand, one conceivable specification would be

\[
p^{i}_{ab}(t) = \frac{p^0_i D^i_{ab}(t)}{S^i_{ab}(t)} = p^0_i \frac{D^i_{ab}(t)}{\min[D^i_{ab}(t), O^i_{ab}(t) V^i_{ab}(t) \rho^0_b \mu_b]} \geq p^0_i,
\] (40)

where \( p^0_i \) are the costs when the supply \( S^i_{ab} \) agrees with the demand \( D^i_{ab} \). Reduced prices would require a slight generalization of equation (29).

Other specifications are, of course, possible as well, as the above formulae partly depend on the strategies of the human decision makers involved.

### 3.6. Calculation of cycle times

Apart from the productivity or throughput \( Q_b \) of a production unit, production managers are also highly interested in the cycle time, i.e. the time interval between the beginning of the generation of a product and its completion. Let us first discuss the process cycle time \( t_b \) between entering the queue of production unit \( b \) and leaving it, assuming that all \( c^i_b \) required educts \( i \) for one production cycle are transported together and located at the same place in the queue. The problem is similar to determining the travel times of vehicles entering a traffic jam [29].

According to equation (22), the average waiting time is determined as the quotient

\[
T_b = L_b / Q_b
\]

of the average queue length \( L_b \) and the throughput \( Q_b \), if production operates in the steady state. Let us now generalize this formula to situations in which the inflow

\[
Q^m_b = (1 - p^*_{k}) \lambda_b
\] (41)

into production unit \( b \) is time dependent and possibly differs from the time-dependent outflow

\[
Q^o_b = (1 - p^*_{k}) \mu_b
\] (42)

(see figure 7). In reality, this time dependence results from fluctuations in the production process, breakdowns of machines, etc. In some cases, one can use the length-dependent formulae

\[
Q^m_b(l_b) = \begin{cases} \lambda_b & \text{if } l_b < S_b \\ 0 & \text{otherwise} \end{cases}
\] (43)
and
\[ Q_b^{\text{out}}(l_b) = \begin{cases} 
\mu_b & \text{if } l_b \geq C_b \\
\mu_b l_b / C_b & \text{if } l_b < C_b 
\end{cases} \]  
(44)

(see section 3.1). We will now derive a delay-differential equation for the cycle time under varying production conditions. For this, let \( l_b(t) \) be the actual length of the queue in production unit \( b \) at time \( t \). The change of this length in time is given by the difference between the inflow and the outflow at time \( t \):
\[ \frac{dl_b}{dt} = Q_b^{\text{in}}(t) - Q_b^{\text{out}}(t). \]  
(45)

If the production unit \( b \) has \( C_b \) channels, an educt entering the production queue at time \( t \) must move forward \( l_b(t) - C_b \) steps before it is finally served by one of the \( C_b \) channels. If, after entering the queue at time \( t \), \( t^*_b(t) \) denotes the waiting time until one of the channels is reached, and if the average processing rate of a single channel is \( \mu_b / C_b \), the average treatment time is given by \( C_b / \mu_b(t + t^*_b) \). The overall time \( t_b \) required for the processing of the product is, therefore, given by the sum of the waiting time \( t^*_b \) and the treatment time by one of the channels:
\[ t_b(t) = t^*_b(t) + \frac{C_b}{\mu_b(t + t^*_b(t))} \]  
(46)

(which replaces the average value \( T_b \). On the other hand, the waiting educts move forward \( Q_b^{\text{out}} \) steps per unit time. For this reason, the waiting time \( t^*_b(t) \) is given by the implicit equation
\[ l_b(t) - C_b = \int_t^{t^*_b(t)} \frac{dl_b}{dt'} Q_b^{\text{out}}(t') = \int_{-\infty}^{t^*_b(t)} dt' Q_b^{\text{out}}(t') - \int_t^{t^*_b(t)} dt' Q_b^{\text{out}}(t'). \]  
(47)

Identifying the time derivative of this equation with equation (45) results in
\[ Q_b^{\text{in}}(t) - Q_b^{\text{out}}(t) = \frac{dl_b}{dt} = Q_b^{\text{out}}(t + t^*_b(t))(1 + \frac{dr^*_b}{dr}) - Q_b^{\text{out}}(t), \]  
(48)

which leads to the delay-differential equation
\[ \frac{dt^*_b}{dr} = \frac{Q_b^{\text{in}}(t)}{Q_b^{\text{out}}(t + t^*_b(t))} - 1. \]  
(49)

As the production initially starts with a waiting time of \( t^*_b(0) = 0 \) (when the factory or production unit \( b \) is opened), this equation can be solved numerically as a function of the outflow \( Q_b^{\text{out}}(t') \). In this way, it is possible to determine the waiting time \( t^*_b(t) \) and process cycle time \( t_b \). In the future, approximation methods will be developed for cases where \( t^*_b(0) \) is not known or \( Q_b^{\text{out}}(t') \) cannot be controlled or anticipated. A rough approximation would be to replace these values by mean values (cf equation (22)).

In a similar way, one can calculate the waiting time \( t^i_b \) in the input buffer \( I^i_b \), resulting in
\[ \frac{dt^i_b}{dt} = \frac{\sum_o S^i_{ob}(t)}{c^i_b Q_b^{\text{out}}(t + t^i_b(t))} - 1. \]  
(50)

The waiting time \( t^j_b \) in the output buffer \( O^j_b \) is given by
\[ \frac{dt^j_b}{dt} = \frac{p^j_b Q_b(t)}{\sum_c S^j_{bc}(t + t^j_b(t))} - 1. \]  
(51)
Finally, the transport time $t_{ab}^t$ between the output buffer of production unit $a$ and the input buffer of production unit $b$ can be estimated by

$$t_{ab}^t = \frac{1}{V_{ab}^i}.$$  \hspace{1cm} (52)

The total cycle time is, then, calculated as the sum of the respective waiting, treatment and transport times.

4. Relation between production processes and macroeconomics

We will now give some support for the ‘macroscopic’ model of supply networks applied in section 2 by relating it to the ‘microscopic’ dynamical production model developed in section 3.

4.1. Just-in-time production

Let us start by deriving a simplified model of production processes. For this, we will summarize the input and output buffers, defining the stock levels (‘inventories’) of markets for the products $i$ by

$$N_i(t) = \sum_b I_i^b(t) + \sum_b O_i^b(t).$$  \hspace{1cm} (53)

According to the balance equations (27) and (28), we find

$$\frac{dN_i}{dt} = \sum_b (p_i^b - c_i^b) Q_b(t).$$  \hspace{1cm} (54)

Moreover, typical for just-in-time production, we will assume negligible input buffers $I_i^b = I_i^0 = 0$, which implies $N_i = \sum_b O_i^b$.  \hspace{1cm} (55)

Consequently, for $\epsilon = 1$ and $V_{ab}^i = V_b^i$, the feeding (arrival) rates become

$$\lambda_b(t) = \rho_0^b \mu_b \min\left(1, \left\{ \frac{V_b^i N_j(t)}{c_b^i} \right\} \right).$$  \hspace{1cm} (56)

Furthermore, for $\delta = 1$ we obtain $N_j^f = \sum_a O_a^j = N_j$ and the related adaptation equations

$$\frac{d(\rho_0^b \mu_b)}{dt} = \frac{1}{\tau_b} \left[ W_b \left( \sum_j \frac{p_b^j X_b^j}{N_j(t)} \right) + \ldots \right] - \rho_0^b \mu_b.$$  \hspace{1cm} (57)

We will now assume $Q_b = \lambda_b$, which is usually given for small enough channel utilizations $\rho_b$ or sufficient storage capacities $S_b$. In that case, one would have to solve equation (57) together with

$$\frac{dN_i}{dt} = \sum_b (p_i^b - c_i^b) \rho_0^b \mu_b \min\left(1, \left\{ \frac{V_b^i N_j(t)}{c_b^i} \right\} \right).$$  \hspace{1cm} (58)

Note that formula (58) maintains non-negative inventories $N_j(t)$, as required. To show this, let us assume that $t$ would be the first point in time where some inventory $N_j$ vanishes, say $N_j(t) = 0$. The inventory $N_j(t)$ could only become negative for $dN_j/dt < 0$, which would require $c_b^j > 0$. But then, the minimum in formula (58) would be $V_b^i N_j(t)/c_b^i = 0$, which contradicts our assumption and proves non-negativity (at least for $V_b^i > 0$).
4.2. Micro–macro link

In the following, we will try to reduce the number of equations by an aggregation procedure, which keeps the structure of the above model equations. Let us first define production sectors \( B \) by summarizing those production units \( b \) which are characterized by the same adaptation times \( \tau_b \) and a proportional throughput \( Q_b(t) \). With suitable constants \( \tau_B \) and \( k_b \), we can then assume

\[
\tau_b = \tau_B \quad \text{and} \quad Q_b(t) = k_b Q_B(t). \tag{59}
\]

The total throughput, processing and feeding rates of production sector \( B \) are defined by

\[
Q_B = \sum_{b \in B} Q_b, \quad \mu_B = \sum_{b \in B} \mu_b, \quad \lambda_B = \sum_{b \in B} \lambda_b, \tag{60}
\]

where \( b \in B \) indicates that \( b \) belongs to production sector \( B \). The equations for the inventories keep their structure

\[
\frac{dN_i}{dt} = \sum_B (p^i_B - c^i_B) Q_B(t), \tag{61}
\]

if we define

\[
p^i_B = \sum_{b \in B} p^i_b k_b \quad \text{and} \quad c^i_B = \sum_{b \in B} c^i_b k_b \tag{62}
\]

with \( k_b = Q_b(0)/Q_B(0) \) and use the relation \( \sum_b = \sum_B \sum_{b \in B} \). Moreover, in order to get the equations

\[
\lambda_B(t) = \rho_0^B \mu_B \min \left( 1, \left\{ \frac{V^j_B N_j(t)}{c^j_B} \right\} \right) \tag{63}
\]

and

\[
\frac{d(\rho_0^B \mu_B)}{dt} = \frac{1}{\tau_B} \left[ W_B \left( \sum_j p^j_B X^j_B N_j(t), \ldots \right) - \rho_0^B \mu_B \right], \tag{64}
\]

we need to set

\[
V^j_B = V^j_B c^j_B / c^j_b \quad \text{and} \quad X^j_B = X^j_B p^j_B / p^j_b \tag{65}
\]

for all \( j \) (requiring that the resulting values on the right-hand sides depend only on \( B \), but not on \( b \in B \)), and

\[
\rho_0^B = \sum_{b \in B} \rho_0^b \mu_b / \mu_B, \quad W_B(1/Z_B) = \sum_{b \in B} W_b(1/Z_b) \tag{66}
\]

with \( 1/Z_B = \sum_j p^j_B X^j_B / N_j = \sum_j p^j_b X^j_b / N_j = 1/Z_b \).

We may also summarize all those markets \( i \) which show a proportional dynamics of \( N_i(t) \) in time. This assumes

\[
N_i(t) = X_i N_i(t), \tag{67}
\]

if we denote the proportionality factors by \( X_i \) and define

\[
N_I(t) = \sum_{i \in I} N_i(t), \tag{68}
\]
where \( i \in I \) indicates that \( i \) is part of the market sector \( I \). Introducing
\[
p_B^I = \sum_{i \in I} p_B^i \quad \text{and} \quad c_B^I = \sum_{i \in I} c_B^i \tag{69}
\]
and summing equation (61) over \( i \in I \) yields the further reduced set of equations
\[
\frac{dN_I}{dt} = \sum_B (p_B^I - c_B^I) Q_B(t). \tag{70}
\]
In order to obtain the equations
\[
\lambda_B(t) = \rho_B^0 \mu_B \min \left( 1, \left\{ \frac{V_B^J N_J(t)}{c_B^J} \right\} \right) \tag{71}
\]
and
\[
\frac{d(\rho_B^0 \mu_B)}{dt} = \frac{1}{\tau_B} \left[ W_B \left( \sum_j p_B^j X_B^j, \ldots \right) - \rho_B^0 \mu_B \right], \tag{72}
\]
we have to define
\[
V_B^J = \min_{j \in J} \left\{ \frac{V_B^J c_B^j X_B^j}{c_B^j} \right\} = \min_{j \in J} \left( \frac{V_B^j c_B^j X_B^j}{c_B^j} \right) \tag{73}
\]
and
\[
X_B^J = \sum_{j \in J} p_B^j X_B^j = \sum_{j \in J} p_B^j X_B^j. \tag{74}
\]
Based on assumptions (59), (65) and (67), one could achieve a considerable reduction in the number of equations, leading from a microscopic model of production processes involving management decisions to a macroscopic model of interacting economic sectors. Conditions (59), (65) and (67) presuppose similar production processes of the summarized production units \( b \) and proportional coefficients regarding the summarized markets \( i \). While our aggregation method can be generalized to time-dependent parameters \( k_b(t) \) and \( X_b(t) \), we still require \( \tau_b = \tau_B \), and the resulting values of \( V_B^j c_B^j / c_B^j \), \( X_B^j p_B^J / p_B^j \) should depend only on \( B \), but not \( b \in B \). These are the main criteria for defining production and market sectors based on empirical data.

4.3. Dynamic input–output model

In equation (70), the sum over \( B \) extends from 1 to \( U + 1 \), with \( B = U + 1 \) representing the final consumer sector. Splitting it up, the resulting equations read
\[
\frac{dN_I}{dt} = \sum_{B=1}^{U} (p_B^I - c_B^I) Q_B(t) - c_{U+1}^I Q_{U+1}(t), \tag{75}
\]
as \( p_{U+1}^I = 0 \). Similarly, \( J \) in formula (71) runs from 0 to \( P \) with \( J = 0 \) representing a market sector for basic resources. Therefore,
\[
\lambda_B(t) = \rho_B^0(t) \mu_B \min \left( 1, \left\{ \frac{V_B^0 N_0(t)}{c_B^0}, \frac{V_B^J N_J(t)}{c_B^J} \right\} \right) \geq 0, \tag{76}
\]
where \( J \in \{1, \ldots, P\} \).

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Let us now concentrate again on the case $Q_B(t) = \lambda_B(t)$ and define the consumption rate from market sector $I$ by

$$Y_I(t) = c^{U}_{U+1} Q_{U+1}(t). \quad (77)$$

When we define the $U$ production sectors $B$ through the respective kind of products they produce, we can set $p^I_B = k_B \delta_{B,I}$ (see figure 9). Without loss of generality, we can choose $k_B = 1$, which just defines the unit quantity in which we measure products of market sector $I$. In the stationary case with $dN_I/dt = 0$ for all $I$, we then obtain the equation

$$\sum_{B=1}^{U} c^{I}_{B} Q_B + Y_I = \sum_{B=1}^{U} p^I_B Q_B = Q_I, \quad (78)$$

which corresponds to Leontief’s input–output model of macroeconomics [24]. Therefore, the above model of supply networks can be considered as a dynamical generalization of this classical economic model.

We should note that, in view of the instability of supply networks, i.e. the existence of the bull-whip effect and of business cycles, the applicability of steady-state concepts in economics is questionable. People have, therefore, tried to formulate dynamical input–output models for a long time in order to take into account investment strategies and other aspects. However, many of these approaches have turned out to be inconsistent or useless (cf [25]). In contrast, the above dynamical input–output model results naturally from a much more general model of supply and production networks. Investment strategies could, for example, be taken into account by a suitable specification of $W_B(\cdots)$, which may be chosen as a function not only of the inventories $N_I$, but also of time derivatives such as $dQ_I/dt$, $dY_I/dt$ or $dN_I/dt$. 

Figure 9. Top: schematic illustration of the actual product flows according to the input–output model with $p^I_B = \delta_{B,I}$. The circle summarizes the production sector (oval) and its market (rectangle). The product flow into its market is represented by a thick arrow, the product flows from markets to production sectors by thin arrows. Bottom: the particular case of a linear supply chain, in which each production unit receives goods only from the previous one.
4.4. Specification of the boundary conditions

In the following, we will discuss some examples. If we assume something like a conservation of materials or value, this implies certain constraints, which reduce the potential complexity of the related supply networks. First of all, the quantity of products $I$ consumed by the production sectors $B$ should be generated somewhere, i.e. 

$$\sum_B c^l_B = \sum_B p^l_B. \quad (79)$$

Second, the quantity of educts consumed by some production sector $B$ corresponds to the quantity of its generated products, i.e. 

$$\sum_I c^l_I = \sum_B p^l_B. \quad (80)$$

For closed systems (with $c^0_B = 0$ and $c^{I+1}_{U+1} = 0$), this implies that the sum over all inventories $N_I$ is constant, i.e. 

$$\sum_I \frac{dN_I}{dt} = \sum_B Q_B(t) \sum_I (p^l_B - c^l_B) = 0. \quad (81)$$

In the following, we will again discuss the case $p^l_B = \delta_{B,I}$. In this particular case, relations (80) and (79) imply the normalization conditions 

$$\sum_B c^l_B = 1 = \sum_I c^l_I. \quad (82)$$

As in section 4.3, we will extend the sums over $I$ from 0 to $U$ and the sums over $B$ from 1 to $U + 1$. In this way, we define 

$$c^0_B = 1 - \sum_I c^l_B \quad \text{and} \quad c^{I+1}_{U+1} = 1 - \sum_B c^l_B. \quad (83)$$

Values $c^0_B > 0$ allow us to describe the inflow of basic resources $I = 0$, while $c^{I+1}_{U+1} > 0$ allows us to describe the depletion of products by a consumer sector $B = U + 1$, the ageing of products, or the generation of products of minor quality. The boundary conditions are completely defined by specifying $N_0(t)$ and $Q_{U+1}(t)$, see equations (75) and (76). This treatment is fully consistent with the intuitive one of linear supply chains in section 2.4.

5. Impact of the supply network’s topology

We will now discuss simulations based on the equations specified in sections 4.1–4.4 for the three different supply networks sketched in figures 10(a)–(c), each with five levels: (a) a linear supply chain with 5 production units, (b) a ‘supply ladder’ with 10 production units and (c) a hierarchical supply tree with 31 production units.

By introducing random variables $\xi^k_J$, which were assumed to be equally distributed in the interval $[-\eta, \eta]$, we have taken into account a heterogeneity $\eta$ in the individual parameters characterizing the different production units. Here, we have chosen $N_0(t) = N_0 = 20 = X^0_B$, $N_I(0) = 20(1 + \xi_I)$, $\tau_B = 180(1 + \xi_B)$, $V^0_B = V = 0.05$, $V^I_B = V c^l_B(1 + \xi^B_B)$ for $J > 0$, $\rho^0_B(0) = W_B(20/N_B(0))$, $Q_{U+1}(t) = W_{U+1}(20/N_0) \min(1, N_I(t)V/c^{I+1}_{U+1})[1 + 0.1 \sin(0.04t)]$ and $A_B = 100/V$.

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Figure 10. Illustration of different supply networks: (a) linear supply chain with five production units, (b) ‘supply ladder’ with five levels and (c) hierarchical supply tree. Circles summarize production sectors and their markets, as in figure 9, while arrows represent the product flows among the sectors.

For the linear supply chain, we have $c^I_B = 1$ if $I$ delivers to $B$, otherwise $c^I_B = 0$:

$$C = (C_{IB}) = (c^I_B) = \begin{pmatrix}
0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0
\end{pmatrix}. \quad (84)$$

For the supply ladder and the hierarchical supply tree, we have $c^I_B = 0.5$ if an arrow points from $I$ to $B$ (see figures 10(b), (c)), otherwise $c^I_B = 0$. Specifically, for the supply ladder we have

$$C = \begin{pmatrix}
0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & \cdots \\
0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & \cdots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & \cdots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \cdots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}, \quad (85)$$
Figure 11. Time-dependent inventories for a supply network with five layers and identical model parameters ($\eta = 0$). The dynamics is the same for all three network topologies displayed in figure 10, i.e. for the linear supply chain, the supply ladder and the hierarchical supply network.

while for the supply hierarchy we have

$$C = \begin{pmatrix}
0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \ldots \\
0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & \ldots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & \ldots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \ldots \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.$$  (86)

$c_B^0$ and $c_{U+1}^I$ are defined in accordance with equation (83). These specifications guarantee that, for $\eta = 0$, i.e. if the production units are characterized by identical parameters, the dynamics of the inventories is the same for all three network topologies discussed (see figure 11). However, the topology matters a great deal if we have a heterogeneity $\eta > 0$ in the model parameters (see figures 12(b), (c)). As our dynamical model of supply networks assumes non-linear interactions, changes of $\eta$ can have large effects. The same applies to small changes in $N_0$ (see figure 6) or in the relaxation times $\tau_B$ (see figure 3).

6. Summary and outlook

In this paper, we have sketched a dynamical theory of supply networks. Here, we could only present first results and indicate possible future research directions, which may contribute to the interdisciplinary field of econophysics [26]. The proposed theory is developed to help understand the complex non-linear phenomena in production and supply networks, in particular their breakdowns, instabilities and inefficiencies. We have also sketched how to derive ‘macroscopic’ equations for the dynamics of a sectorally structured economics from ‘microscopic’ equations describing single production processes, involving strategic decisions of production managers reflected by the control functions $W_b$. The resulting model is a dynamical generalization of the classical input–output model of macroeconomics. For the particular case of a linear supply chain,
Figure 12. Time-dependent inventories for supply networks in the case of heterogeneous parameters (with a heterogeneity of $\eta = 0.2$). The dynamics drastically depends on the topology of the supply network: (a) linear supply chain, (b) supply ladder, (c) hierarchical supply network. Compared with figure 11 one can conclude that heterogeneity in supply networks can considerably decrease the undesired oscillation amplitudes in the inventories. The strongest effect by far is found for supply ladders, which is relevant for the design of robust supply networks.
one can relate it to the Hilliges–Weidlich model, which was originally developed for traffic flow. These equations can describe the ‘bull-whip effect’ due to their linear instability in a certain regime of operation. The underlying mechanism is a slow adaptation of the processing rate to changes in the order flows or stock levels in the market. Interestingly, the resulting oscillations in the inventories of the different products have a characteristic frequency which can be much lower than the underlying fast variations in the consumption rate. Depending on the network structure, these oscillations are synchronized among different economic sectors and may explain business cycles as a self-organized phenomenon with slow dynamics. These conclusions are not restricted to linear supply chains, but can be generalized to many other supply networks which are linearly unstable with respect to perturbations. In reality, business cycles are, of course, less regular and of smaller amplitudes than in figures 5 or 11. However, the above model allows us to reflect these aspects in a natural way by inclusion of fluctuations, heterogeneity, additional capacity constraints and realistic network structures. Investigations with empirical data and for particular kinds of networks are on the way. It should also be noted that already deterministic models of supply networks can show irregular, non-periodic behaviour such as chaotic dynamics [15, 27], which calls for suitable control concepts [28].

Compared with traffic dynamics, economic and production systems have some interesting new features: instead of a continuous space, we have discrete production units or sectors, and the production speed as a function of the inventories is different from the empirical velocity–density relation in traffic flow. Due to the minimum condition (56), production systems may operate in different regimes, and small changes of parameters may have tremendous effects. For example, we may have a transition from small oscillations of relatively high frequency to large oscillations of low frequency. Apart from this, the management strategies can vary to a large extent, and with this the control functions $W_b(\cdots)$. With suitable strategies, the oscillations can be mitigated or even suppressed [9, 17, 18]. Moreover, production systems are frequently supply networks with complex topologies rather than linear supply chains, i.e. they have additional features compared with (more or less) one-dimensional freeway traffic. They are more comparable with street networks of cities [29].

Apart from some equations which were not further applied in this study, most of the proposed model equations were conservation equations, equations given by the product flows, or relations derived from stochastic concepts used in queueing theory. They reflect the transport and interaction of products, so that the physics of driven many-particle systems and of complex systems can make some significant contributions to the new multidisciplinary field of self-organization phenomena in production and supply networks.

Future work will have to address questions such as the relevance of the network structure to the resulting dynamics, possible control strategies, the role of the market and pricing mechanism, etc. This particularly concerns the specification of the function $W_b$, and the equations suggested in the subsection on delivery networks and price dynamics. In this paper, we have already seen that the supply network’s topology and the level of heterogeneity in production systems have a significant impact on the resulting dynamics: parameter changes can have tremendous effects. For example, a heterogeneity in the parameters characterizing the different production units can stabilize the production and the market considerably, while in traffic flow, heterogeneity normally has a destabilization effect. The stabilization of supply networks through heterogeneity probably could explain why the variations in economic systems appear to be less dramatic than in our simulations, but additional inefficiencies and capacity restrictions (corresponding to finite values of $S_b$, $I_{b,\text{max}}$ and $O_{b,\text{max}}$) are probably another reason. This and the role of heterogeneity for the micro-
macro link will be studied in more detail in the future. Should it turn out that the micro–macro link requires a high degree of homogeneity in the parameters, it would be preferable to simulate economic dynamics based on a microscopic model of production and supply networks in the future. The tendency for globalization certainly increases the degree of homogeneity, but this also tends to generate larger oscillations in the inventories, i.e. more serious over- and underproduction. At least in some markets, there are definite signs of a development in this direction. Forthcoming publications will, therefore, investigate alternative control strategies (including forecasts). First results on how to decrease the instability of supply chains can be found in [9, 17, 18].

6.1. Advantages, extensions and potential applications

As the variables in our dynamical model of supply and production networks are operational and measurable, the model can be tested and calibrated with empirical data. Moreover, it is flexible and easy to generalize. For this reason, it can be adapted to various applications. Our approach can be related to microscopic considerations such as queueing theory or event-driven (Monte Carlo) simulations of production processes, but, since it focuses on the average dynamics, it is numerically much more efficient and, therefore, suitable for online control. Nevertheless, the formulae can be extended by noise terms to reflect stochastic effects. Our system of coupled differential equations would then become a coupled system of stochastic differential equations (Langevin equations), where the noise amplitudes would be determined via relationships from queueing theory.

The dynamical theory of supply chains appears to be a promising field with many research opportunities. It is not just useful for a deeper understanding of the origin and dynamics of business cycles or for the optimization of production processes and supply networks. It could also contribute to the further improvement of existing traffic control strategies in urban street networks or to the development of more robust routing algorithms for internet traffic. Apart from this, the model can be viewed as a dynamic multi-player game, where the individual control functions \( W_b \) reflect the strategies of the players \( b \) in terms of quantities \( N_i \), which represent the information feedback available to them. Generalizing this idea, the above model of supply networks may serve as a basis for particular kinds of neural networks. We are now setting up different projects in these directions, and cooperation is very welcome.

Acknowledgments

The author is grateful for inspiring discussions with Arne Kesting, Megan Khoshyaran, Christian Kühnert, Stefan Lämmer, Tadeusz Płatkowski, Dick Sanders, Thomas Seidel, Torsten Werner and Ulrich Witt. He would also like to thank Tilo Grigat for preparing the illustrations.

References

[1] Nagatani T 2002 The physics of traffic jams Rep. Prog. Phys. 65 1331
Chowdhury D, Santen L and Schadschneider A 2000 Statistical physics of vehicular traffic and some related systems Phys. Rep. 329 199
Helbing D 2001 Traffic and related self-driven many-particle systems Rev. Mod. Phys. 73 1067
[2] Santra S B, Schwarzer S and Herrmann H 1996 Fluid-induced particle segregation in sheared granular assemblies Phys. Rev. E 54 5066

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
[3] Dzubiella J and Löwen H 2002 Pattern formation in driven colloidal mixtures: tilted driving forces and re-entrant crystal freezing J. Phys.: Condens. Matter 14 9383
Dzubiella J, Hoffmann G P and Löwen H 2002 Lane formation in colloidal mixtures driven by an external field Phys. Rev. E 65 021402

[4] Helbing D and Molnár P 1995 Social force model of pedestrian dynamics Phys. Rev. E 51 4282
Helbing D and Vicsek T 1999 Optimal self-organization New J. Phys. 1 13.1
Couzin I D and Franks N R 2002 Self-organized lane formation and optimized traffic flow in army ants Proc. R. Soc. B 02PB0606.1

[5] Helbing D, Hennecke A and Treiber M 1999 Phase diagram of traffic states in the presence of inhomogeneities Phys. Rev. Lett. 82 4360

[6] Helbing D, Farkas I and Vicsek T 2000 Simulating dynamical features of escape panic Nature 407 487

[7] To K, Lai P-Y and Pak H K 2001 Jamming of granular flow in a two-dimensional hopper Phys. Rev. Lett. 86 71

[8] Kerner B S and Konhäuser P 1994 Structure and parameters of clusters in traffic flow Phys. Rev. E 50 54

[9] Daganzo C 2003 A Theory of Supply Chains (New York: Springer)

[10] Marthaler D, Armbruster D and Ringhofer C 2002 A mesoscopic approach to the simulation of semiconductor supply chains Proc. Int. Conf. on Modeling and Analysis of Semiconductor Manufacturing ed G Mackulak et al p 365
Armbruster D, Marthaler D and Ringhofer C 2003 Modeling a re-entrant factory Oper. Res. submitted
Armbruster D, Marthaler D and Ringhofer C 2002 Kinetic and fluid model hierarchies for supply chains SIAM J. Multiscale Modeling Simulation submitted
Rem B and Armbruster D 2003 Control and synchronization in switched arrival systems Chaos at press
Diaz-Rivera I, Armbruster D and Taylor T 2000 Periodic orbits in a class of re-entrant manufacturing systems Math. Oper. Res. 25 708

[11] Witt U and Sun G-Z 2002 Myopic behavior and cycles in aggregate output Jahrbücher f. Nationalökonomie u. Statistik vol 222/3 (Stuttgart: Lucius & Lucius) p 366

[12] Schumpeter J A 1934 The Theory of Economic Development (Cambridge, MA: Harvard University Press) Samuelson P 1939 Interations between the multiplier analysis and principle of acceleration Rev. Econ. Stud. 21 75
Kaldor N 1940 A model of the trade cycle Econ. J. 50 78

[13] Helbing D 2003 Modeling supply chains and business cycles as unstable transport phenomena Preprint cond-mat/0301204

[14] Hilliges M and Weidlich W 1995 A phenomenological model for dynamic traffic flow in networks Transp. Res. B 29 407
Hilliges M 1995 Ein phänomenologisches Modell des dynamischen Verkehrsflosses in Schnellstraßennetzen (Aachen: Shaker)

[15] Sterman J D 2000 Business Dynamics (Boston, MA: McGraw-Hill)

[16] Mosekilde E and Larsen E R 1988 Syst. Dyn. Rev. 4 131–47

[17] Nagatani T and Helbing D 2003 Stabilization of a linear supply chain Preprint cond-mat/0304476

[18] Helbing D and Šeba P 2003 Stability of regular and random supply networks Working paper (unpublished)

[19] Bando M, Hasebe K, Nakayama A, Shibata A and Sugiyama Y 1995 Dynamical model of traffic congestion and numerical simulation Phys. Rev. E 51 1035
Bando M, Hasebe K, Nakanishi K, Nakayama A, Shibata A and Sugiyama Y 1995 Phenomenological study of dynamical model of traffic flow J. Physique 5 1389
Bando M, Hasebe K, Nakayama A, Shibata A and Sugiyama Y 1994 Structure stability of congestion in traffic dynamics Jpn. J. Ind. Appl. Math. 11 203

[20] Kerner B S, Klenov S L and Konhäuser P 1997 Asymptotic theory of traffic jams Phys. Rev. E 56 4200

New Journal of Physics 5 (2003) 90.1–90.28 (http://www.njp.org/)
Herrmann M and Kerner B S 1998 Local cluster effect in different traffic flow models Physica A 255 163
[21] Treiber M, Hennecke A and Helbing D 1999 Derivation, properties, and simulation of a gas-kinetic-based, non-local traffic model Phys. Rev. E 59 239
[22] Hall R 1991 Queueing Methods for Service and Manufacturing (Upper Saddle River, NJ: Prentice Hall)
Saaty T 1983 Elements of Queueing Theory with Applications (New York: Dover)
Fischer K and Hertel G 1990 Bedienungsprozesse im Transportwesen: Grundlagen und Anwendungen der Bedienungstheorie (Berlin: Transpress)
Gross D and Harris C M 1998 Fundamentals of Queueing Theory (New York: Wiley)
Bose S K 2001 Introduction to Queueing Systems (New York: Kluwer Academic)
Cox D R and Smith W L 1991 Queues (Boca Raton, FL: Chemical Rubber Company Press)
Ayres J B 2000 Handbook of Supply Chain Management (Boca Raton, FL: St. Lucie)
Nahmias S 2000 Production and Operations Analysis (Chicago: McGraw-Hill)
Hopp W J and Spearman M L 2000 Factory Physics (Boston: McGraw-Hill)
[23] Feistel R and Ebeling W 1989 Evolution of Complex Systems (Dordrecht: Kluwer Academic)
[24] Leontief W W 1966 Input–Output Economics (New York: Oxford University Press)
[25] Holub H-W and Schnabl H 1994 Input–Output-Analyse (Munich: Oldenbourg)
Meyer U and Schumann J 1977 Das dynamische Input–Output-Modell als Modell gleichgewichtigen Wachstums Z. gesamte Staatswiss. 133 1
[26] Mantegna R N and Stanley H E 1999 Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge: Cambridge University Press)
Bouchaud J-P and Potters M 2000 Theory of Financial Risk: From Statistical Physics to Risk Management (Cambridge: Cambridge University Press)
Levy H, Levy M and Solomon S 2000 Microscopic Simulation of Financial Markets (San Diego, CA: Academic)
Challet D and Zhang Y-C 1998 On the minority game: analytical and numerical studies Physica A 256 514
[27] Beaumariage T and Kempf K 1994 The nature and origin of chaos in manufacturing systems Proc. Conf. 5th IEEE/SEMI Advanced Semiconductor Manufacturing (Piscataway, NJ: IEEE) pp 169–74
Ushio T, Ueda H and Hirai K 1995 Controlling chaos in a switched arrival system Syst. Control Lett. 26 335–9
Katzorke I and Pikovsky A 2000 Chaos and complexity in simple models of production dynamics Discrete Dyn. Nat. Soc. 5 179–87
[28] Holyst J A, Hagel T, Haag G and Weidlich W 1996 How to control a chaotic economy? J. Evol. Econ. 6 31–42
Bunimovich L A 1999 Controlling production lines Handbook of Chaos Control ed H Schuster (Berlin: Wiley–VCH) pp 324–43
Bartholdi J J and Eisenstein D D 1996 A production line that balances itself Oper. Res. 44 21–34
Zavadlav E, McClain J O and Thomas L J 1996 Self-buffering, self-balancing, self-flushing production lines Manage. Sci. 42 1151–64
[29] Helbing D 2003 A section-based queueing-theoretical traffic model for congestion and travel time analysis J. Phys. A: Math. Gen. submitted