

Some remarks on stability and robustness of production networks based on fluid models

Bernd Scholz-Reiter¹, Fabian Wirth², Sergey Dashkovsky³, Michael Schönlein², Thomas Makuschewitz¹, Michael Kosmykov³

¹ Planning and Control of Production Systems (PSPS), BIBA – Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Bremen, Germany

² Institute of Mathematics, University of Würzburg, Würzburg, Germany

³ Centre of Industrial Mathematics, University of Bremen, Bremen, Germany

Abstract The dynamics of complex, large-scale production networks present an important issue not only for the management of such networks but also for scientific research. This problem is usually investigated either by numerical simulations or by mathematical analysis of the associated queueing model. One major property of stable production networks is the robustness of these networks with respect to perturbations of the arrival process of jobs from outside. Given a generic structure of the considered network with several production locations, different products and re-entrant material flows the determination of robustness is non trivial. In this paper we use a fluid model approach to analyse the robustness of queueing networks. First conditions for stability of a fluid network are introduced. These conditions allow to assess the dynamic behaviour of the production network. Second the obtained results are investigated with the help of a simulation of the fluid and queueing network. Simulations of a test case scenario accompany the results of the theoretical analysis.

1 Introduction

Modern production networks often consist of production facilities which are geographically distributed around the globe [1]. The generic structure of these networks leads to an increased complexity, which together with the dynamics of modern production networks poses a unique challenge for management and research [2]. In particular perturbations or uncertainties of the market requirements may lead to an unstable behaviour of a network [3] and the network may not be able to fulfil the customer demand. This raises the question, what type or size of perturbations is admissible without destroying the stability of a production network? This question can be answered using the stability radius of a network,

which reflects the magnitude of the smallest possible perturbation that destabilizes the system [4].

The dynamics of such complex production networks can be modelled by multi-class queueing networks. In [5, 6] Dai presented a new approach to investigate the stability of such networks using a fluid model, which is an analogue deterministic continuous model of the discrete stochastic model. The stability of a corresponding fluid limit model implies the stability of the original queueing network. In comparison to a queueing model the stability of a fluid model can be determined more easily. Based on the stability of a fluid model we present an approach to determine the stability radius. The results of this approach are validated using simulations of both the fluid model and the corresponding queueing network.

The outline of the paper is as follows. Section 2 briefly introduces the notation of a fluid model and the necessary and sufficient conditions of stability. This introduction is accompanied by the necessary adaptations for modelling a production network as a fluid model. In order to illustrate the characteristic dynamic behaviour of a fluid and queueing network we introduce in section 3 a test scenario. This scenario is used in section 4 to investigate the robustness of the underlying production network. Section 5 closes with some conclusions and an outlook to future research.

2 Description of the fluid model

We follow the model description from [7]. The considered network consists of locations \mathcal{S}_j with $j \in \mathcal{J} = \{1, 2, \dots, J\}$ and different types of products \mathcal{P}_k with $k \in \mathcal{K} = \{1, 2, \dots, K\}$. Every type of product is processed exclusively at one location. So there is a many-to-one mapping $\sigma : \{\mathcal{P}_1, \dots, \mathcal{P}_K\} \rightarrow \{\mathcal{S}_1, \dots, \mathcal{S}_J\}$. The mapping σ generates the so called constituency matrix C , with $c_{jk} = 1$ if $\sigma(\mathcal{P}_k) = \mathcal{S}_j$ or $c_{jk} = 0$ otherwise. For every location the set $\mathcal{C}(\mathcal{S}_j) := \{\mathcal{P}_k \in \{\mathcal{P}_1, \dots, \mathcal{P}_K\} : \sigma(\mathcal{P}_k) = \mathcal{S}_j\}$ is assumed to be nonempty. Further every type of product \mathcal{P}_k has an exogenous arrival rate α_k and a process rate μ_k of products per time unit. After a product of type \mathcal{P}_k was processed at location $\sigma(\mathcal{P}_k)$ it has the possibility either to leave the network or to become a product of type \mathcal{P}_l . The transition matrix P sets the proportion of processed products that either turn to their next location or leave the network. To be precise, p_{kl} denotes the proportion of products of type \mathcal{P}_k that become products of type \mathcal{P}_l upon service completion. Hence $1 - \sum_{l=1}^K p_{kl}$ is the proportion that eventually leaves the network. It is assumed that the $K \times K$ matrix P has spectral radius strictly less than one, i.e. all products leave the network. The initial amount of products is represented through the K dimensional vector $Q(0)$. The model of the network is given by (α, μ, P, C) and $Q(0)$. The performance is described by the K dimensional product level process $\{Q(t) : t \geq 0\}$ and the K dimensional allocation process $\{T(t) : t \geq 0\}$, where $Q_k(t)$ denotes the amount of products \mathcal{P}_k in the network at time t and $T_k(t)$ denotes the total amount of time

in the interval $[0, t]$ that location $\sigma(\mathcal{P}_k)$ has devoted to processing products of type \mathcal{P}_k . For brevity we omit the calligraphic letters in the subscript. The next step is to fix a policy that predetermines the order in which the arriving products are processed at each location. We use the so called priority discipline, i.e. we consider a permutation $\pi : \{1, \dots, K\} \rightarrow \{1, \dots, K\}$. Given two types of products \mathcal{P}_k and \mathcal{P}_l that are processed at the same location $\sigma(\mathcal{P}_k) = \sigma(\mathcal{P}_l)$, we say that \mathcal{P}_l has higher priority than \mathcal{P}_k if $\pi(l) < \pi(k)$. So products of type \mathcal{P}_k are not processed as long as there are products of type \mathcal{P}_l . For \mathcal{P}_k the set

$$H_k = \{ \mathcal{P}_l : \mathcal{P}_l \in C(\sigma(\mathcal{P}_k)), \pi(\mathcal{P}_l) \leq \pi(\mathcal{P}_k) \}$$

denotes all products \mathcal{P}_l that have a higher priority than \mathcal{P}_k . Finally the process of unused capacity $Y = \{Y(t) : t \geq 0\}$ is introduced, where $Y_k(t)$ denotes the cumulative remaining capacity of location $\sigma(\mathcal{P}_k)$ for processing products of types that have strictly lower priority than products of type \mathcal{P}_k . The dynamics of the network under priority discipline are summarized as follows

$$Q(t) = Q(0) + \alpha t - (I - P^T)MT(t) \geq 0, \quad (1)$$

$$T(0) = 0 \text{ and } T(t) \text{ is nondecreasing} \quad (2)$$

$$Y_k(t) = t - \sum_{l \in H_k} T_l(t) \text{ is nondecreasing, } k \in \mathcal{K} \quad (3)$$

$$0 = \int_0^\infty Q_k(t) dY_k(t), \quad k \in \mathcal{K} \quad (4)$$

where $M = \text{diag}(\mu_1, \dots, \mu_K)$. Equation (4) describes the work-conserving property of the network and relation (1) is called the flow balance relation. The work-conserving property means that the idle time for a product of type \mathcal{P}_k increases if and only if $Q_k(t) = 0$, i.e. there is no product of type \mathcal{P}_k in the network waiting for being processed. The following Theorem [8] guarantees the existence of such a work-conserving allocation process.

Theorem 1: For any fluid network (α, μ, P, C) with $Q(0)$ there is at least one work-conserving allocation T .

Let n_j denote the number of product types being processed at location \mathcal{S}_l . For $1 \leq l \leq m \leq n_j$ let the product type \mathcal{P}_l have higher priority than \mathcal{P}_m . The allocation process $T(t)$ for the priority discipline is assumed to be non pre-emptive. That is, if products of type \mathcal{P}_m are processed at a given time, then this production is completed regardless of the arrival of other types. Any pair $(Q(t), T(t))$ that satisfies (1) – (4) is called a fluid solution of the work-conserving fluid network. The set of all feasible fluid level processes is denoted as

$$\Phi = \{Q(t) : \text{there is a } T(t) \text{ such that } (Q(t), T(t)) \text{ is a fluid solution}\}.$$

For $x \in \mathbb{R}^K$ we use $\|x\|_1 = \sum_{k=1}^K |x_k|$.

Definition 1: A fluid network is said to be stable, if there exists a time $\tau > 0$

such that $Q(\tau + \cdot) \equiv 0$ for any $Q(\cdot) \in \Phi$ with $\|Q(0)\|_1 = 1$.

We cite the following necessary condition for the stability of a fluid network [8]. Here $<$ has to be understood componentwise.

Theorem 2: *If a fluid network (α, μ, P, C) is stable, it holds that*

$$C M^{-1} (I - P^T)^{-1} \alpha < e, \quad (5)$$

where $e = (1, \dots, 1)^T$.

Remark 1: Common notations in the literature are $\lambda := (I - P^T)\alpha$ and $\rho := C M^{-1}\lambda$ which denote the total arrival rate and the nominal workload. Hence the condition (5) can be stated as $\rho_j < 1$ for all $j \in \mathcal{J}$.

Further in [8] there is also a sufficient condition for stability. Here a symmetric matrix A is called strictly copositive, if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$ with $x_i \geq 0$ and $x^T A x = 0$ only if $x = 0$. Here $x^- := \min \{0, x\}$.

Theorem 3: *A fluid network (α, μ, P, C) is stable if there exists a $K \times K$ symmetric strictly copositive matrix A such that, for $k = 1, \dots, K$*

$$\sum_{i=1}^K \alpha_i a_{ik} - \min_{\mathcal{P}_i \in \mathcal{C}(\mathcal{S}_k)} h_{ik} - \sum_{j=1, \mathcal{S}_j \neq \sigma(\mathcal{P}_k)}^J \left(\min_{\mathcal{P}_i \in \mathcal{C}(\mathcal{S}_j)} h_{ik} \right)^- < 0,$$

where $H = M(I - P)A$.

Adaption to production networks

The fluid model introduced in Section 2 can be used to model and to analyse the dynamics of a production network. To this end the production facilities of suppliers and OEM, warehouses and distribution centres are modelled as locations. These locations are numbered from 1 to J . The intermediate and final products of the network are classified into product types 1 to K that are processed or serviced by the locations. Thus the structure of the production network is given by the constituency and the transition matrix. The external and internal dynamics of the considered network are captured by the inflow rates of the product types and the processing rates of the product types at the assigned locations. The transition matrix of the fluid model contains the fragmentation of the material flows within the production network. The initial work in progress of the product types is given by $Q(0)$. If a location serves more than on product type a priority rule for the service needs to be determined.

3 Test scenario

In this paper we simulate a network with three types of products and two locations. The parameters for the test scenario are given below.

$$\alpha = \begin{pmatrix} 0.15 \\ 0.15 \\ 0.10 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0.6 \\ 0.9 \\ 0.5 \end{pmatrix}, \quad P = \begin{pmatrix} 0.25 & 0.15 & 0.20 \\ 0.05 & 0.25 & 0.15 \\ 0.20 & 0.25 & 0.10 \end{pmatrix}$$

$$Q(0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

A schematic illustration is given in Figure 1.

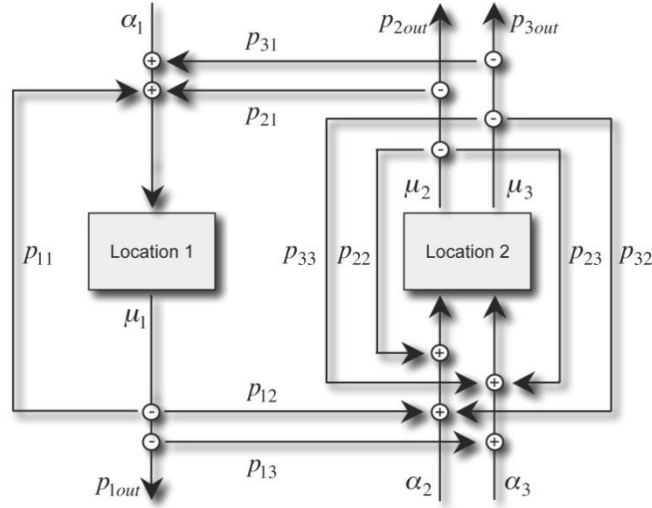


Fig. 1: Fluid network with two locations processing three types of products.

The priorities at the second location are chosen such that products of type 3 have higher priority than the products of type 2, i.e. $\pi = (\{1\}, \{3,2\})$. Figure 2 shows the behaviour of the work in progress (WIP) under non pre-emptive priority discipline, i.e. the WIP of type 2 products increases as long as the WIP of product type 3 has not reached zero. At the time $Q_3(t) = 0$ location 2 is processing type 2 products until the WIP reaches zero. Meanwhile the WIP of type 3 products increases again. Since the configuration is stable the repeated increase of the WIP of product types 2 and 3 becomes smaller and smaller and thus the frequency of the allocation changes at the second station is becoming higher. Consequently the WIP of all three product types reaches zero at some time and there is a time τ such that the WIP stay zero beyond τ . Figure 3 shows the WIP of the associated queueing network. Since the fluid model is stable the WIP of the queueing network becomes zero from time to time.

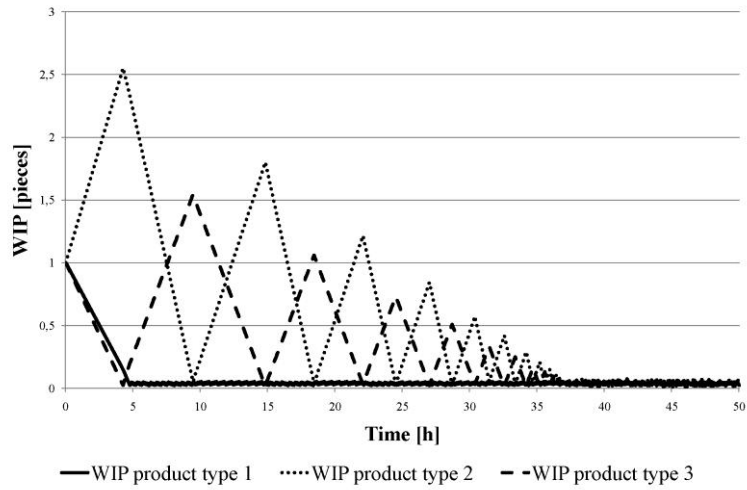


Fig. 2: WIP of the fluid model for the production network.

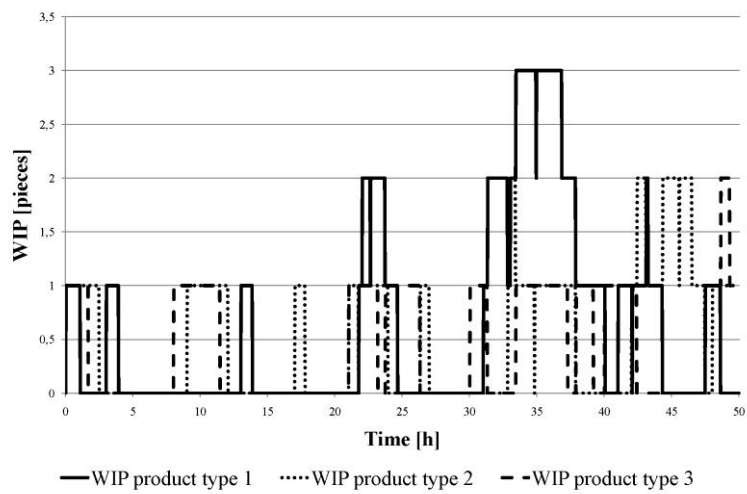


Fig. 3: WIP of the queueing model for the production network.

4 An approach to describe robustness of production networks

In this section we focus on the question of quantifying the size of admissible perturbations of the external arrival rate. The minimal size of a destabilizing perturbation will be called the stability radius of the network. If the condition (5) is violated Theorem 2 states that the network cannot be stable. The set of all feasible arrival rates, i.e. all $\alpha \in \mathbb{R}^K$ such that that condition (5) is satisfied, forms a polyhedron. So we increase the arrival rate by introducing a vector $\delta \in \mathbb{R}^K$ and consider a perturbation of the following form $\alpha \rightsquigarrow \alpha_\delta = \alpha + \delta$. Then we consider the fluid network $(\alpha_\delta, \mu, P, C)$ and use the following notation $\Delta = C M^{-1} (I - P^T)^{-1}$. The quantity of interest is the radius $\|\delta\|_1$ of smallest ball such that stability will no longer hold. Thus we give the following Definition.

Definition 2: *The stability radius of the network (α, μ, P, C) is*

$$r(\alpha, \mu, P, C) = \inf \{ \|\delta\|_1 : (\alpha_\delta, \mu, P, C) \text{ is not stable} \}.$$

Remark 2: By Theorem 2 an upper bound for the stability radius is given by

$$\inf \{ \|\delta\|_1 : \Delta(\alpha_\delta) \not\prec e \}.$$

In the following we give a geometric interpretation of the previous remark. In Figure 4 the light grey area represents the set of all arrival rates that satisfy condition (5). Let α be an interior point.

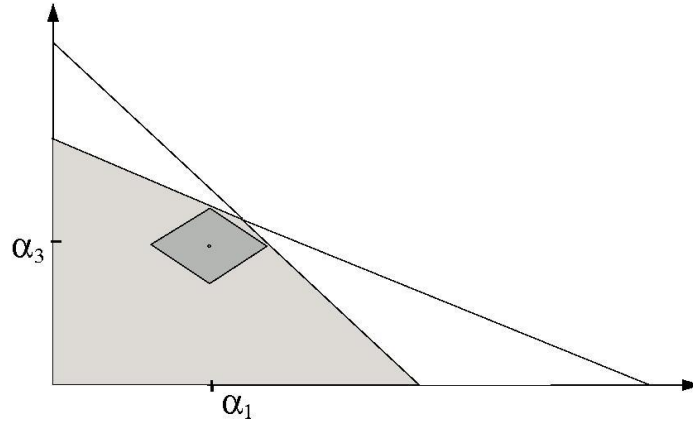


Fig. 4: Illustration of the stability radius in the case where the second arrival rate is not perturbed.

The dark grey area is the largest open norm ball $B_r(\alpha)$ around α that lies completely in the light grey domain. The upper bound for the stability radius is calculated as follows, where we use $\Delta_j \cdot \alpha_\delta$ to denote the scalar product of the j -th row with α_δ .

$$d_1 := \min \|\delta\|_1 \text{ such that } \Delta_1 \cdot \alpha_\delta \geq 1$$

$$d_2 := \min \|\delta\|_1 \text{ such that } \Delta_2 \cdot \alpha_\delta \geq 1$$

$$r(\alpha, \mu, P, C) \leq \min_{j=1,2} d_j$$

In regard to the test scenario Figure 2 shows that the WIP of product type 1 reaches and remains zero long before the WIP of products type 2 and 3 do. Intuitively, one might expect that the network is able to cope with an additional arrival rate of type 1 products. A calculation for the test scenario using Remark 2 leads to $r(\alpha, \mu, P, C) \leq 0.0547781$ and the corresponding vector of perturbation is $\delta = (0, 0, 0.0547781)^T$. Since this is only an upper bound for the stability radius there may be arrival rates that are inside the dark grey area which lead to instability. But this subset is very small as the following shows. For the perturbation $\delta = (0, 0, 0.05428493)^T$ the matrix

$$A = \begin{pmatrix} 10 & 0.374 & 0.53 \\ 0.374 & 0.645 & 0.916 \\ 0.53 & 0.916 & 1.3 \end{pmatrix}$$

satisfies the sufficient condition of Theorem 3 and the network is stable. The corresponding perturbed arrival rate is $\alpha_\delta = (0.15, 0.15, 0.15428493)^T$. In order to present the product level process and the queue length process clearly the following figures are generated for the perturbed arrival rate $\alpha_\delta = (0.15, 0.15, 0.153)^T$.

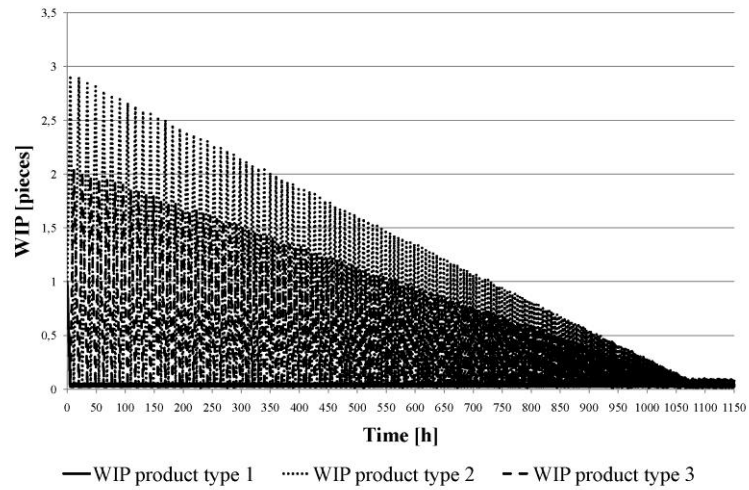


Fig. 5: WIP of the fluid model with perturbed arrival rate α_δ .

Figures 5 and 6 show the WIP of the fluid and queueing model for the perturbed production network. Further the pictures show that the time until the WIP of all types of products reaches zero has increased enormously as well as that the amplitude of the WIP rises by a factor of ten.

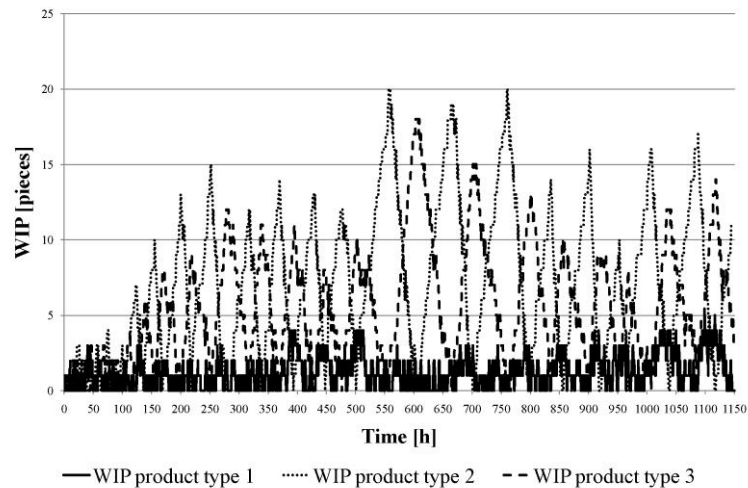


Fig. 6: WIP of the queueing network with perturbed arrival rate α_δ .

5 Conclusions and outlook

We have introduced a fluid model that can be applied to a generic structure of locations in a production network. In particular, a scenario for two locations processing three types of products has been concerned. The parameter setting was chosen such that the fluid model is stable. Figure 3 illustrates the behaviour of the corresponding queueing network. In the subsequent section the stability radius has been defined, which gives an upper bound to the additional arrival rates δ such that the production network is still able to handle the arrival rate α_δ . Again Figure 6 illustrates this for the arrival rate α_δ . From Section 4 it can be seen that for the queueing network model corresponding to α_δ the amplitude of the queue length increases by a factor ten.

In future research we will also consider perturbations of the production rates μ as well as perturbations of the transition matrix P . Moreover it is aspired to obtain analytical estimations of the queue length compared to increasing arrival rates or decreasing production rates.

Acknowledgement: This research was funded by Volkswagen Foundation.

References

- [1] Wiendahl, H-P., Lutz, S. (2002): Production in networks, *Annals of the CIRP*, 51(2), 1-14
- [2] Scholz-Reiter, B., Wirth, F., Dashkovskiy, S., Jagalski, T., Makuschewitz, T. (2008): Analyse der Dynamik großskaliger Netzwerke in der Logistik. *Industrie Management*, 3, 37-40.
- [3] Scholz-Reiter, B., Wirth, F., Freitag, M., Dashkovskiy, S., Jagalski, T., de Beer, C., Ruffer, B. (2005): Some remarks on the stability of production networks, In: *Proceedings of the International Scientific Annual Conference on Operations Research*. Springer, Bremen, Germany, 91-96
- [4] Hinrichsen, P., Pritchard, J. (2005): *Mathematical Systems Theory I*. Springer, Berlin
- [5] Dai, J. (1995): On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models, *Ann. Appl. Probab.*, 5, 49-77
- [6] Dai, J., Jennings, O. (2003): Stability of general processing networks, In: *Stochastic modeling and optimization* (Eds: D. Yao, H. Zhang, X. Zhou), Springer, New York
- [7] Ye, H.Q., Chen, H. (2001): Lyapunov method for the stability of fluid networks, *Operations Research Letters* 28, 125-136
- [8] Chen, H. (1995): Fluid approximations and stability of multiclass queueing networks: work-conserving disciplines, *Ann. Appl. Probab.*, 5, 637-655