Instability in Stochastic and Fluid Queueing Networks

John J. Hasenbein

University of Texas at Austin

Joint work with David Gamarnik, IBM

Stability via Fluid Models

Fluid Model Stable

$$\Downarrow$$

Queueing Model Stable

Stability via Fluid Models

Fluid Model Stable

$$\Uparrow \qquad\qquad \Downarrow$$

Queueing Model Stable

The Fluid Model Program

Stability: given a particular policy or class of policies, under what conditions on the system parameters is a queueing network stable?

- Extract the mean value fluid model under the scheduling policy
- Analyze the fluid solutions
- Use the set of fluid solutions to determine stability of the stochastic model

Can an exact stability analysis be achieved with the fluid model program?

The Fluid Model Program

- We can answer the question with a firm

The Fluid Model Program

- We can answer the question with a firm . . . yes and no.

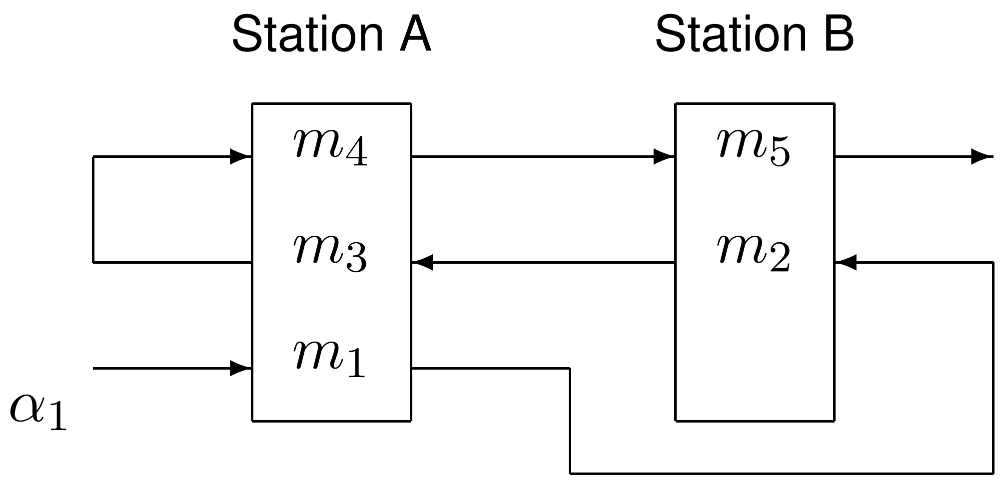
The Fluid Model Program

- We can answer the question with a firm . . . yes and no.
- **NO** - Under an arbitrary but fixed policy (Dai, H, and Vandevate 2004).

The Fluid Model Program

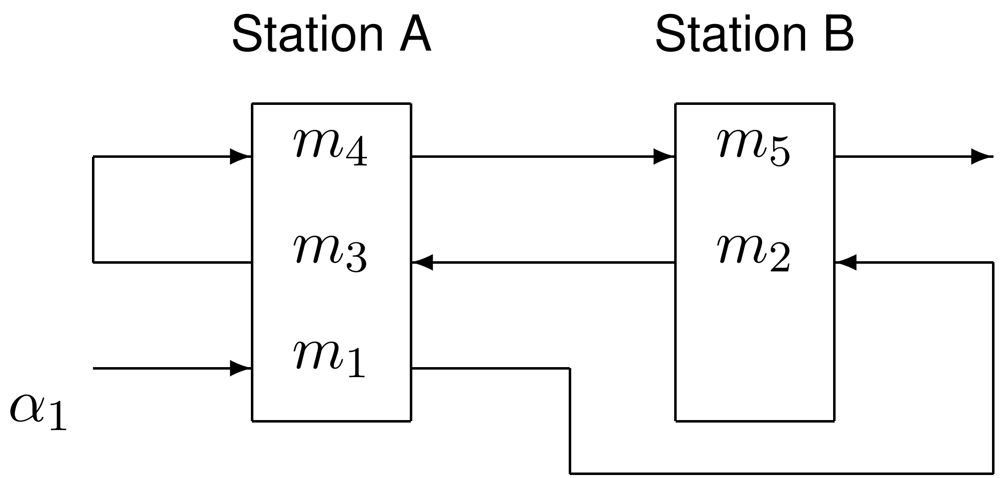
- We can answer the question with a firm . . . yes and no.
- **NO** - Under an arbitrary but fixed policy (Dai, H, and Vandevate 2004).
- **YES** - For the class of non-idling policies (Gamarnik and H 2004).

Multiclass Queueing Networks



- A queueing network of single-server workstations.
- Stochastic customer arrivals from the outside with a exogenous rate α_k.
- The average processing time for a customer at stage i is m_i. The processing *rate* at stage i is $\mu_i = 1/m_i$.

Mean-Value Fluid Networks



- Set of dynamical equations, which contain only α_1, $\vec{m}$ from the stochastic network.
- Examine set of fluid solutions $\{Q(t), T(t)\}$.
- Stability analysis via fluid solutions.

Non-Idling Mean Value Fluid Model

We are interested in solutions to the fluid model equations:

$$\bar{Q}(t) = \bar{Q}(0) + \alpha t - (I - P')\bar{D}(t),$$

$$\bar{D}(t) = \mathrm{diag}(\mu)\bar{T}(t),$$

$$\bar{W}(t) = C\mathrm{diag}(m)\bar{Q}(t),$$

$\bar{T}(0) = 0,\ \bar{T}(\cdot)$ is non-decreasing,

$\bar{Y}_j(t) = t - \sum_{k \in C(j)} \bar{T}_k(t),$ non-decreasing

$\bar{Y}_j(t)$ can increase only when $\bar{W}_j(t) = 0$

$$\bar{Q}(t) \geq 0,$$

The Fluid Model

- A solution $\{(\bar{Q}(t), \bar{T}(t)), t \geq 0\}$ to the above equations is called a *non-idling fluid solution*.
- $\mathcal{F}(\mathrm{lim}) \subset \mathcal{F}(\mathrm{eq})$, set of fluid solutions.
- Fluid solution need not be a fluid limit. ↶

Stochastic Stability

- Let $\{X(t), t \geq 0\}$ be the state process of the queueing network.
- **Definition 1:** The queueing network is said to be *stable* if $\{X(t), t \geq 0\}$ is positive Harris recurrent.
- If the network is stable when operating under all non-idling policies, then it is said to be *globally stable*.

Stochastic Stability

- **Definition 2:** A reentrant line is said to be *rate stable* if starting from any initial state x,

$$\mathbb{P}_x\left\{\lim_{t\to\infty}\frac{D_k(t)}{t}=\alpha_1\right\}=1,$$

  where $D_k(t)$ is the number of jobs which have departed buffer k in $[0,t]$.
- If the network is rate stable under all non-idling policies then it is *globally rate stable*.

A Stability Theorem

- **Theorem 1 – Strong Stability** (Dai 95, Stolyar 95)
  - If the fluid model is stable, then the queueing network is stable.
- The fluid model is stable if, there exists a T such that for every fluid solution with $||\bar{Q}(0)|| \leq 1$, $\bar{Q}(t) = 0$ for all $t \geq T$.
- Note: fluid model is not stable if just one solution does not go to zero.
- Is this “bad” fluid solution bad enough?

Chen's Stability Theorem

- **Theorem 2 – Weak Stability** (Chen 95)
  - If the fluid model is weakly stable, then the queueing network is rate stable.
- A fluid model is *weakly stable* if for all fluid solutions with $\bar{Q}(0) = 0$, $\bar{Q}(t) = 0$ for all $t \geq 0$. ↷
- Note: fluid model is not rate stable if just one solution pops up from zero.
- Is this "bad" fluid solution bad enough?

A Full Converse to Chen's Result

- **Theorem 3 – Converse for Two Stations** (Gamarnik and H 04)
  - For two station networks: if the fluid model is not globally weakly stable, then the queueing network is not globally rate stable.
- If there exists a fluid solution solution that "pops up" from zero, then there exists a non-idling scheduling policy under which $Q(t) \to \infty$ *linearly* a.s.
- Stochastic primitives must satisfy some large deviations bounds.
- Theorem 3 holds for a network of any size if it satisfies a Finite Decomposition Property.

Large Deviations Assumptions

- Let $\{Z_n, n \geq 1\}$ be an i.i.d sequence with $E\mathbf{Z}_1 = \alpha$.
- For every $\epsilon > 0$ there exist constants $L = L(\epsilon), V = V(\epsilon) > 0$ such that for any $z > 0$

$$\mathbb{P}\left(\Big| \sum_{1 \leq i \leq n} \mathbf{Z}_i - z - \alpha n \Big| \geq \epsilon n \;\Big|\; \mathbf{Z}_1 \geq z\right) \leq V e^{-Ln},$$

for all $n \geq 1$.
- The counting process $\mathbf{N}(t) \equiv \max\{n : \mathbf{Z}_1 + \cdots + \mathbf{Z}_n \leq t\}$ satisfies

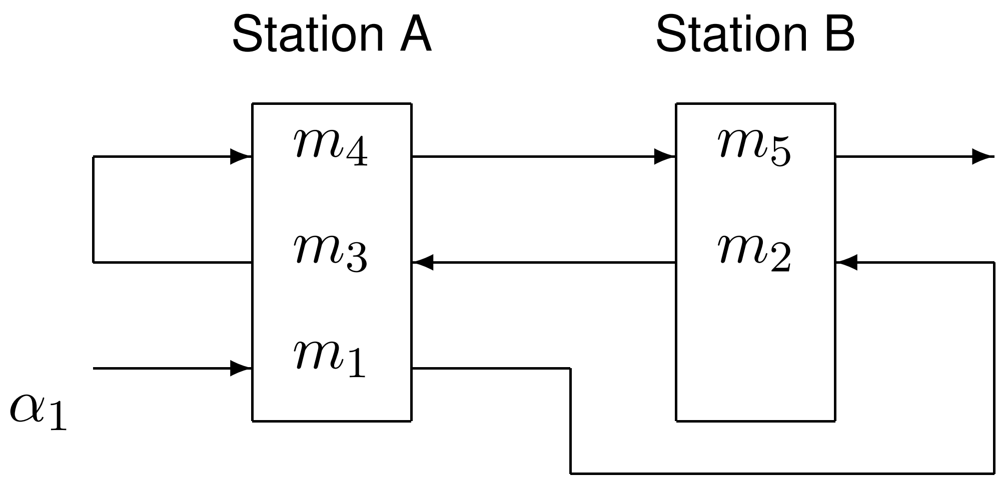
$$\mathbb{P}\left(\Big|\mathbf{N}(t+z) - \frac{t}{\alpha}\Big| \geq \epsilon t \;\Big|\; \mathbf{Z}_1 \geq z\right) \leq V e^{-Lt},$$

for all $t \geq 0$.

Corollary to Theorem 3

- **Corollary:** two station network is globally rate stable if and only if the virtual station and pushstart conditions of Dai and VandeVate hold.
- Dai and VandeVate proved *sufficiency* of the conditions.
- The necessity of the conditions, for networks with "pushstarts" was an open question.
- First "full converse" to fluid stability theorems.

Corollary to Theorem 3



Any stochastic network with the topology above is globally rate stable iff:

$$\rho_1 = \alpha_1(m_1 + m_3 + m_4) \leq 1,$$

$$\rho_2 = \alpha_1(m_2 + m_5) \leq 1,$$

$$\rho_{\text{ps}} = \alpha_1 \left(\frac{m_3}{1 - \alpha_1 m_1} + m_5 \right) \leq 1.$$

Necessary and sufficient conditions for stochastic stability.

Other Converses

- **Meyn 95:** if all fluid limits eventually diverge at some uniform rate $\Rightarrow$ unstable
- **Dai 96:** if all fluid limits are “weakly unstable” $\Rightarrow$ unstable.
- **Puhalskii & Rybko 00:** if there exists a set of “close” fluid limits which satisfy a uniform divergence condition, $\Rightarrow$ not positive Harris recurrent.
- **Meyn 04:** if there is a set of fluid limits which satisfy a uniform homogeneity condition $\Rightarrow$ unstable
- All transience results require demonstration of some sort of “uniform divergence” for a **set** of fluid limits.

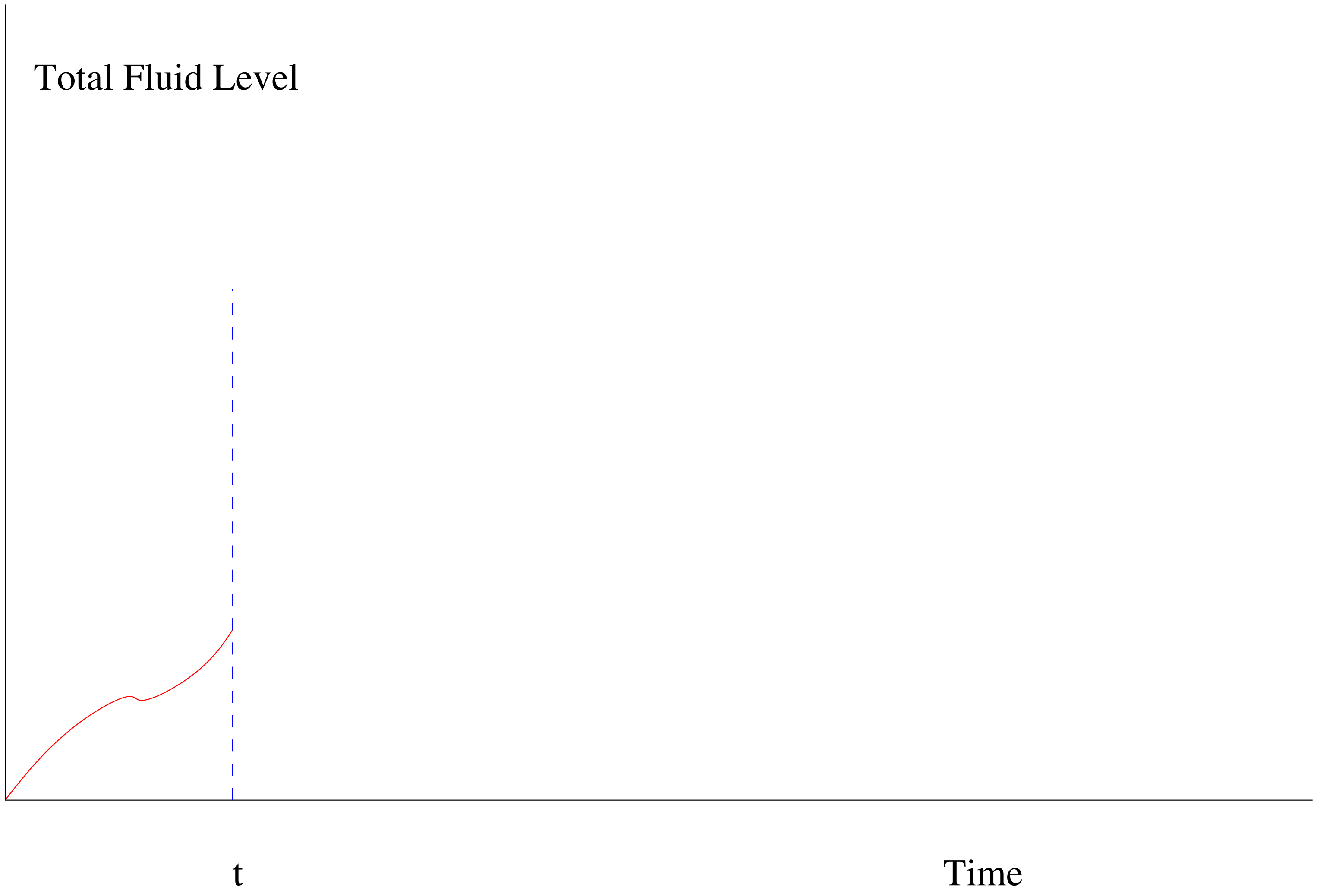
Proof Outline

- If there exists a fluid solution with $\bar{Q}(0) = 0$ and $\bar{Q}(t_0) > 0$ then there exist solutions for which

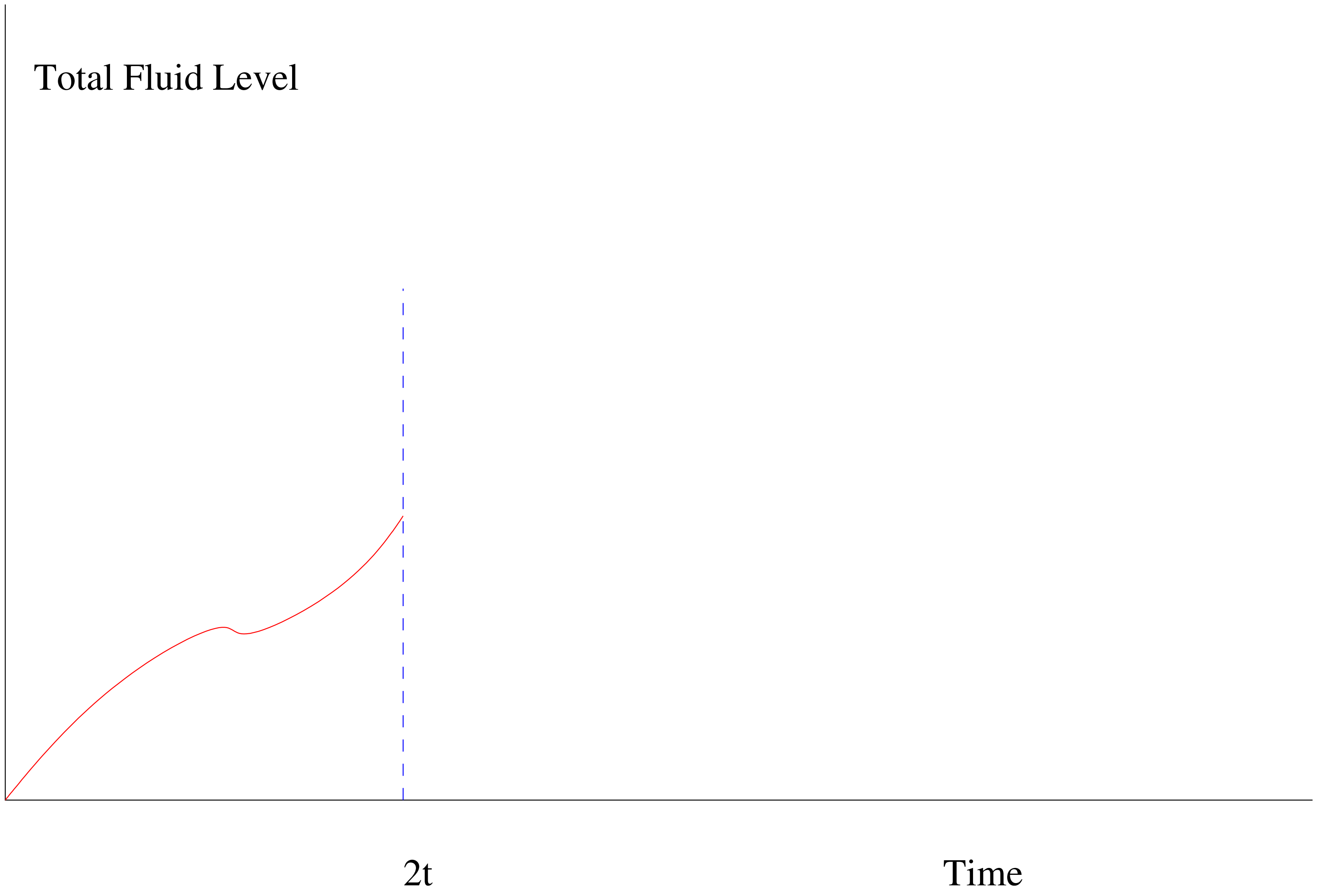
  $$\liminf_{t \to \infty} \frac{||\bar{Q}(t)||}{t} > 0$$

  from any initial point $\bar{Q}(0) = q$.
- In other words, there exist linearly divergent solutions.
- Propose a non-idling scheduling policy which attempts to "follow" the fluid model.
- Using large deviations bounds for processes, show that the stochastic paths will follow the fluid solution with "high" probability.

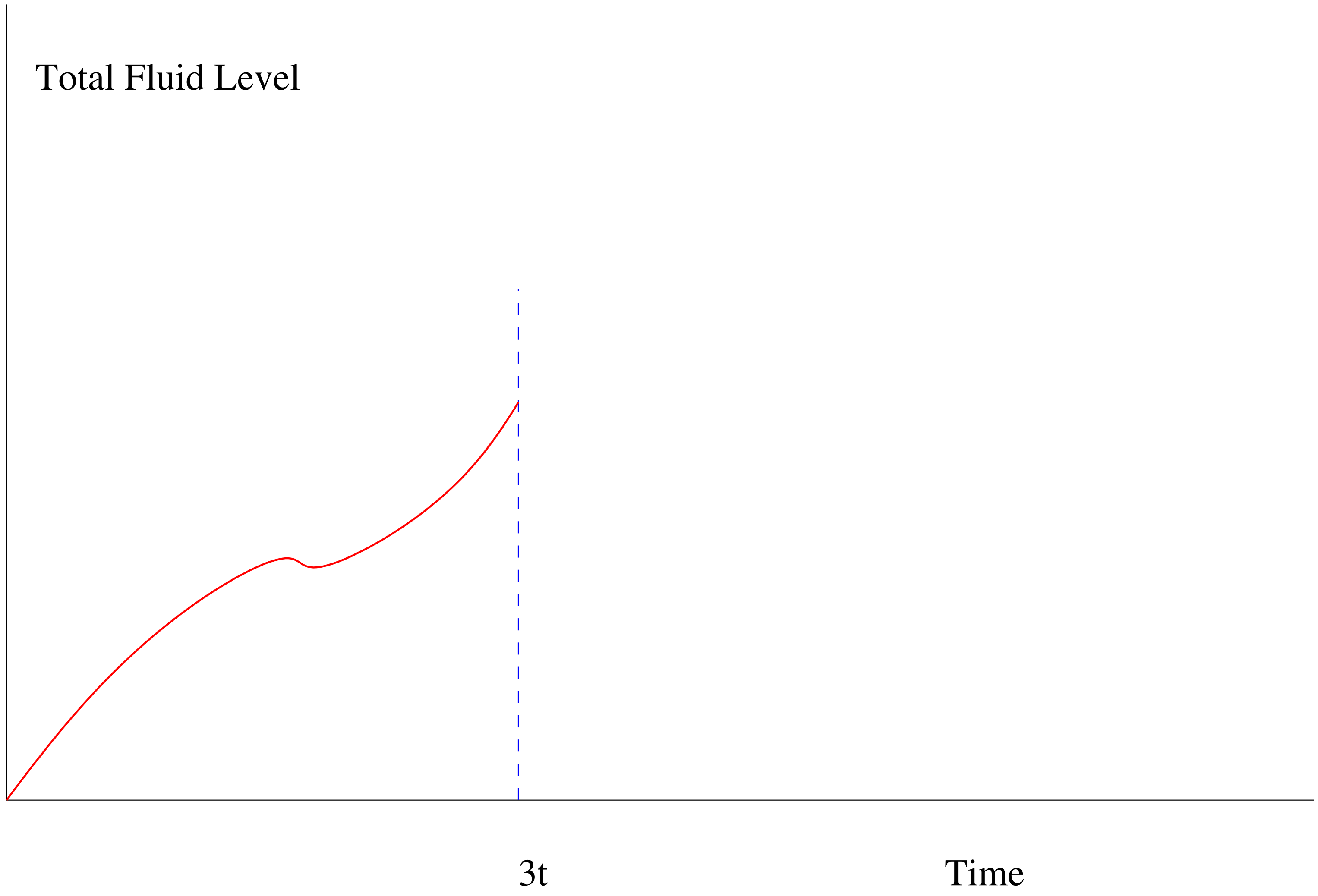
Weakly Unstable to Divergent



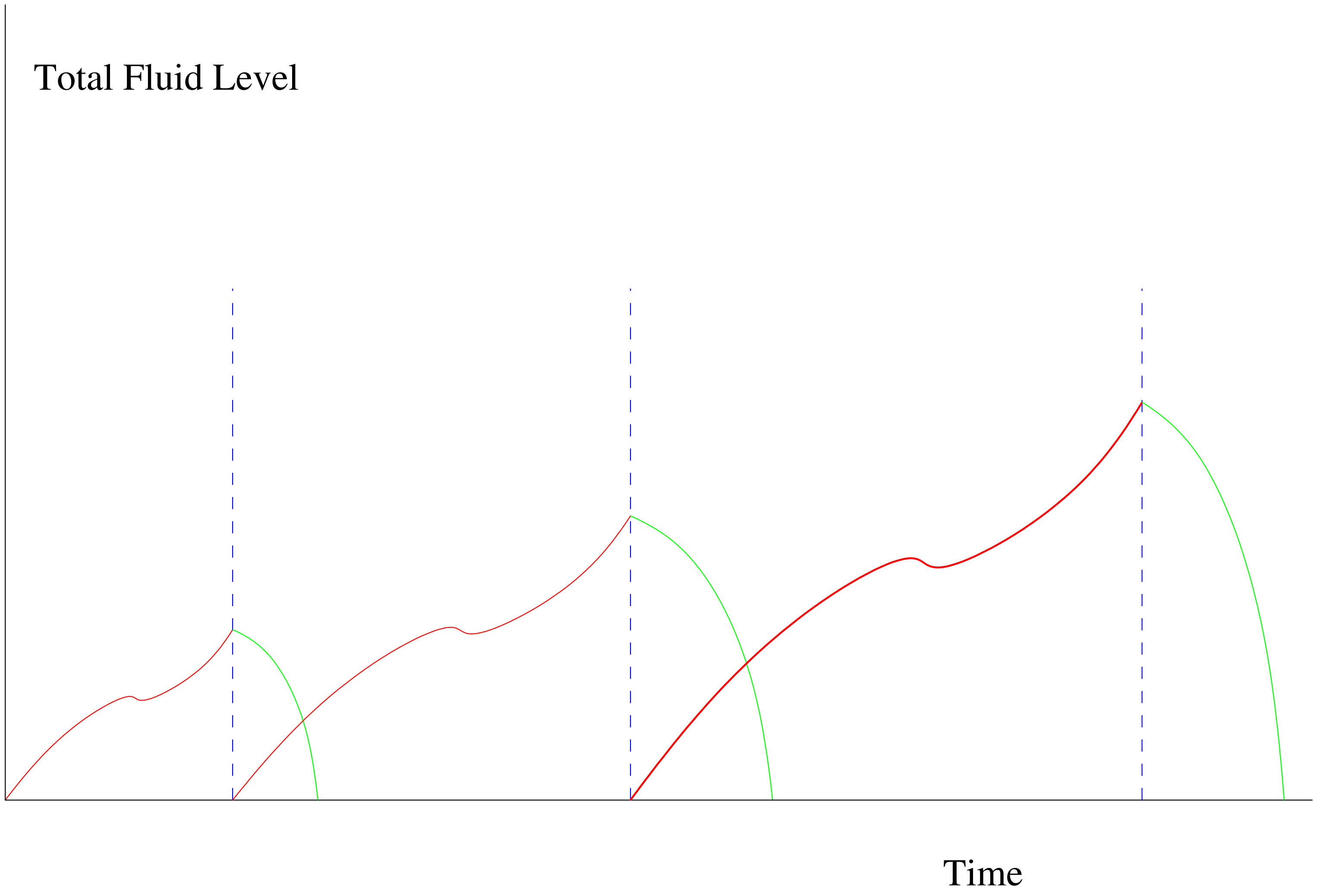
Weakly Unstable to Divergent



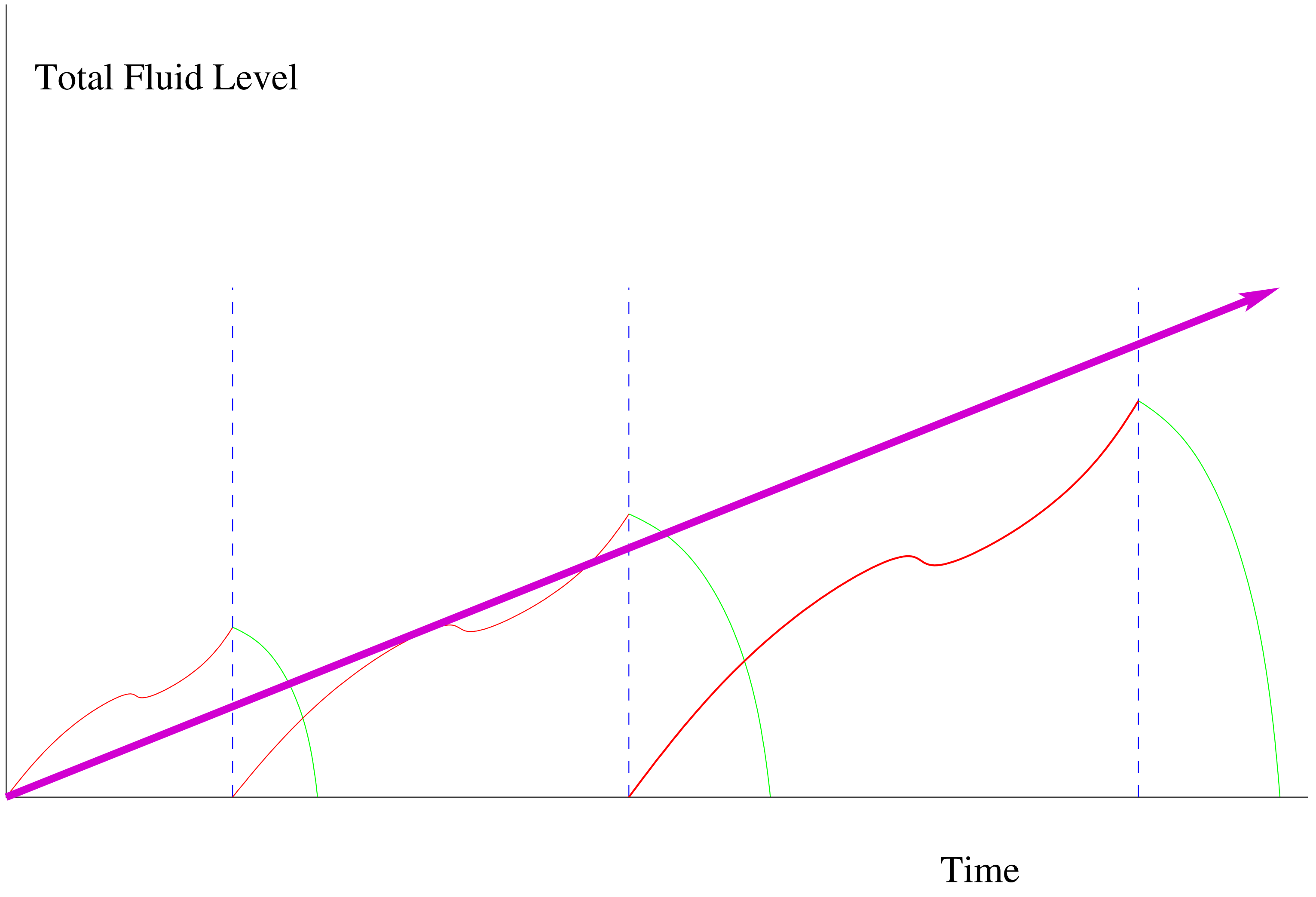
Weakly Unstable to Divergent



Weakly Unstable to Divergent



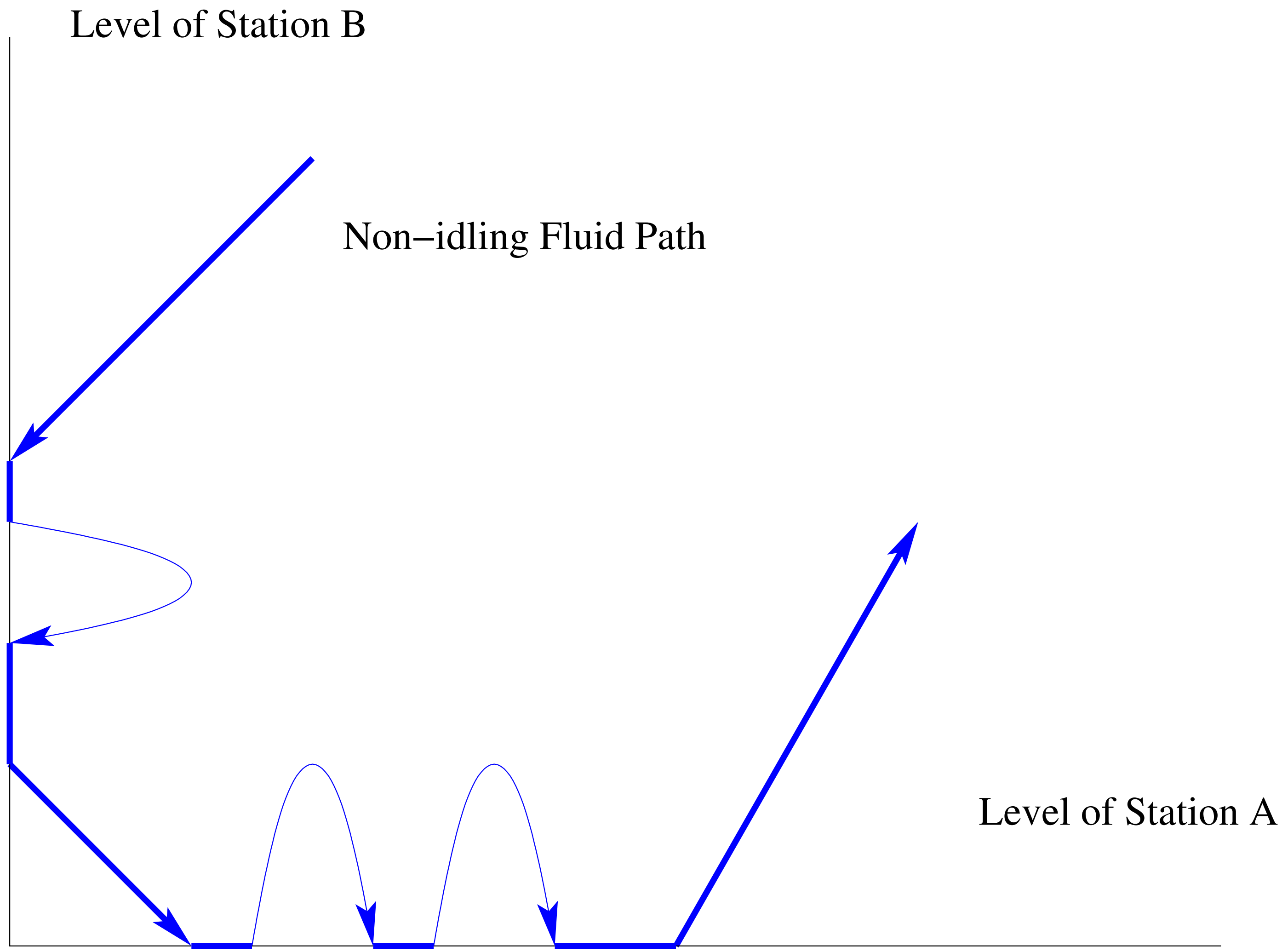
Weakly Unstable to Divergent



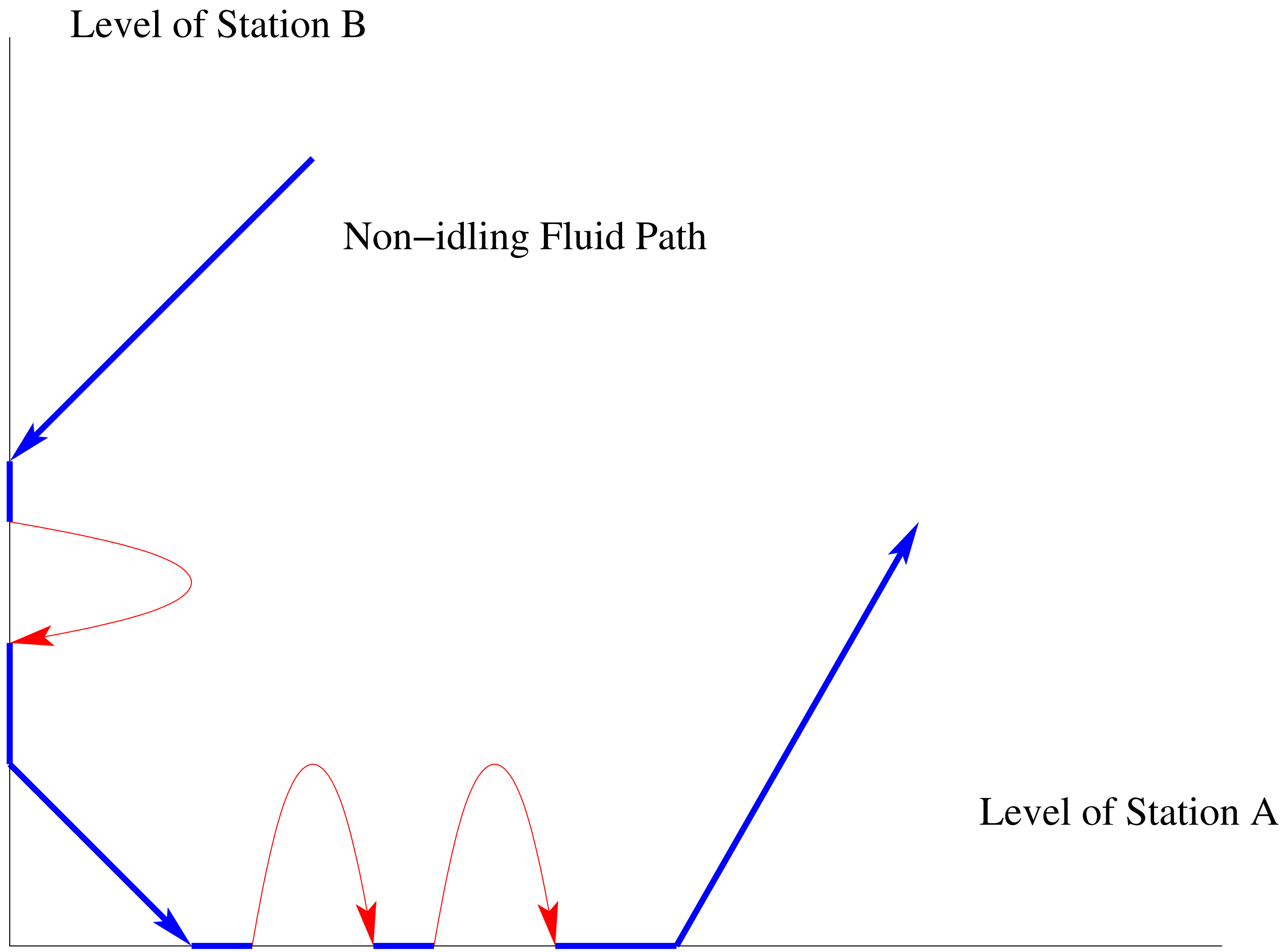
Why only two station networks?

- Crucial part of large deviations proof - Fluid Decomposition Property of fluid models.
  - Consider any non-idling solution with $||\bar{Q}(t)|| > 0$ for $t \in [0, T]$.
  - Then there exists another non-idling solution and times $0 = t_0, t_1, \ldots t_n = T$ such that on each interval for each station j, $\bar{W}_j(t) > 0$ for all $t \in (t_i, t_{i+1})$ or $\bar{W}_j(t) = 0$ for all $t \in (t_i, t_{i+1})$.
  - Finite number of intervals. Stations are either busy or empty for entire interval.
- **Theorem 4:** FDP holds for all fluid models arising from two station queueing networks.

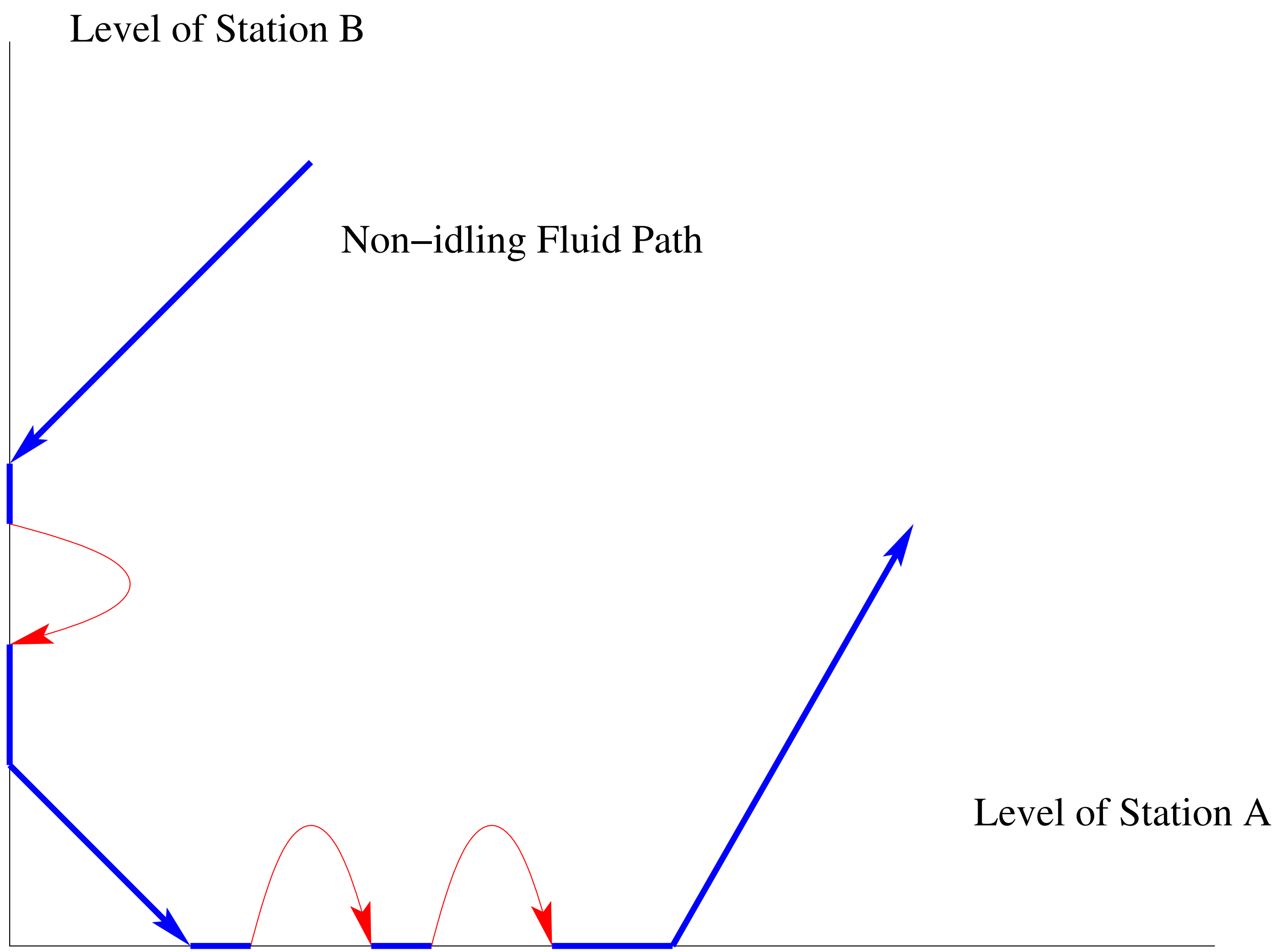
Proof of FDP in 2D



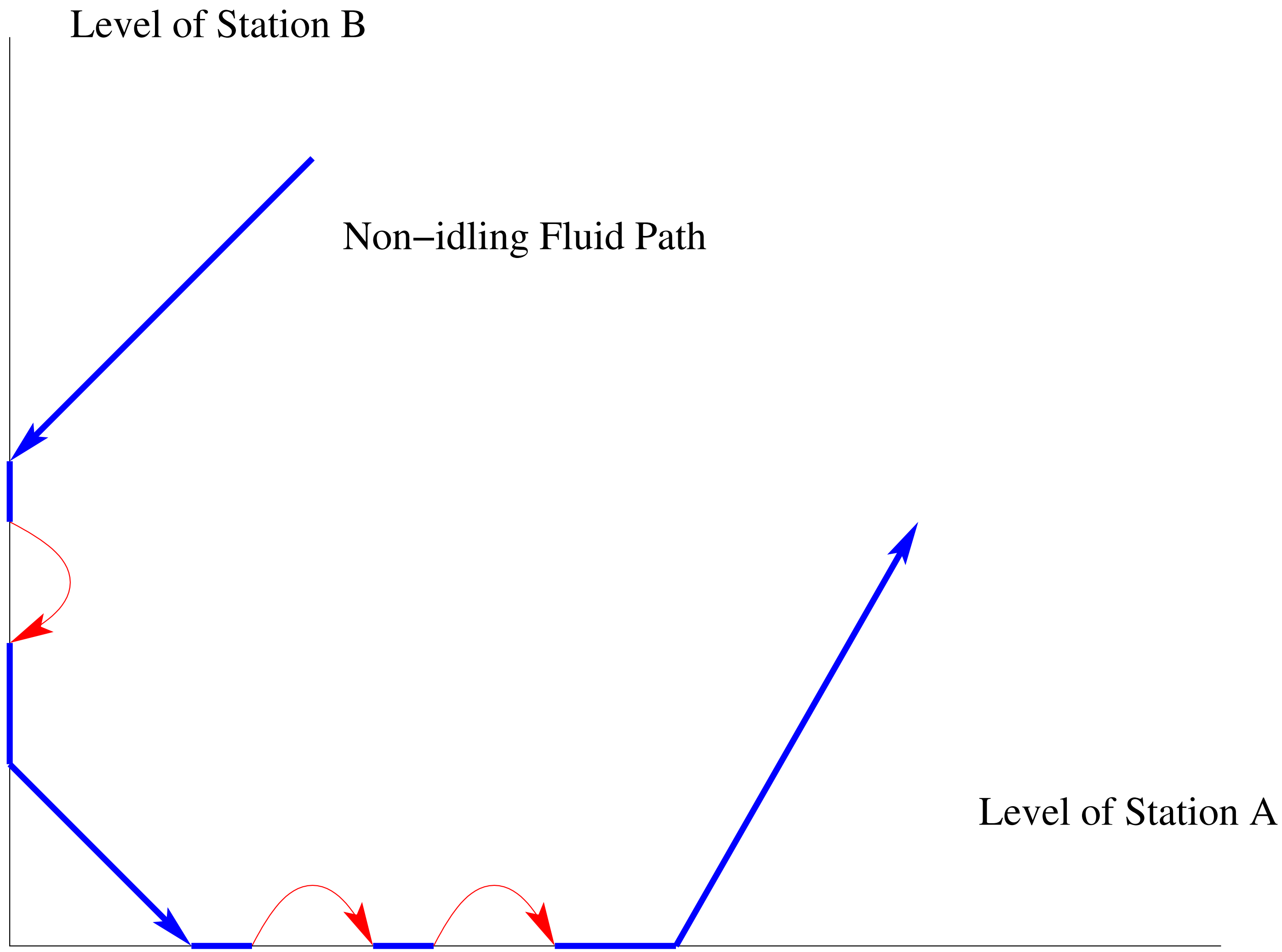
Proof of FDP in 2D



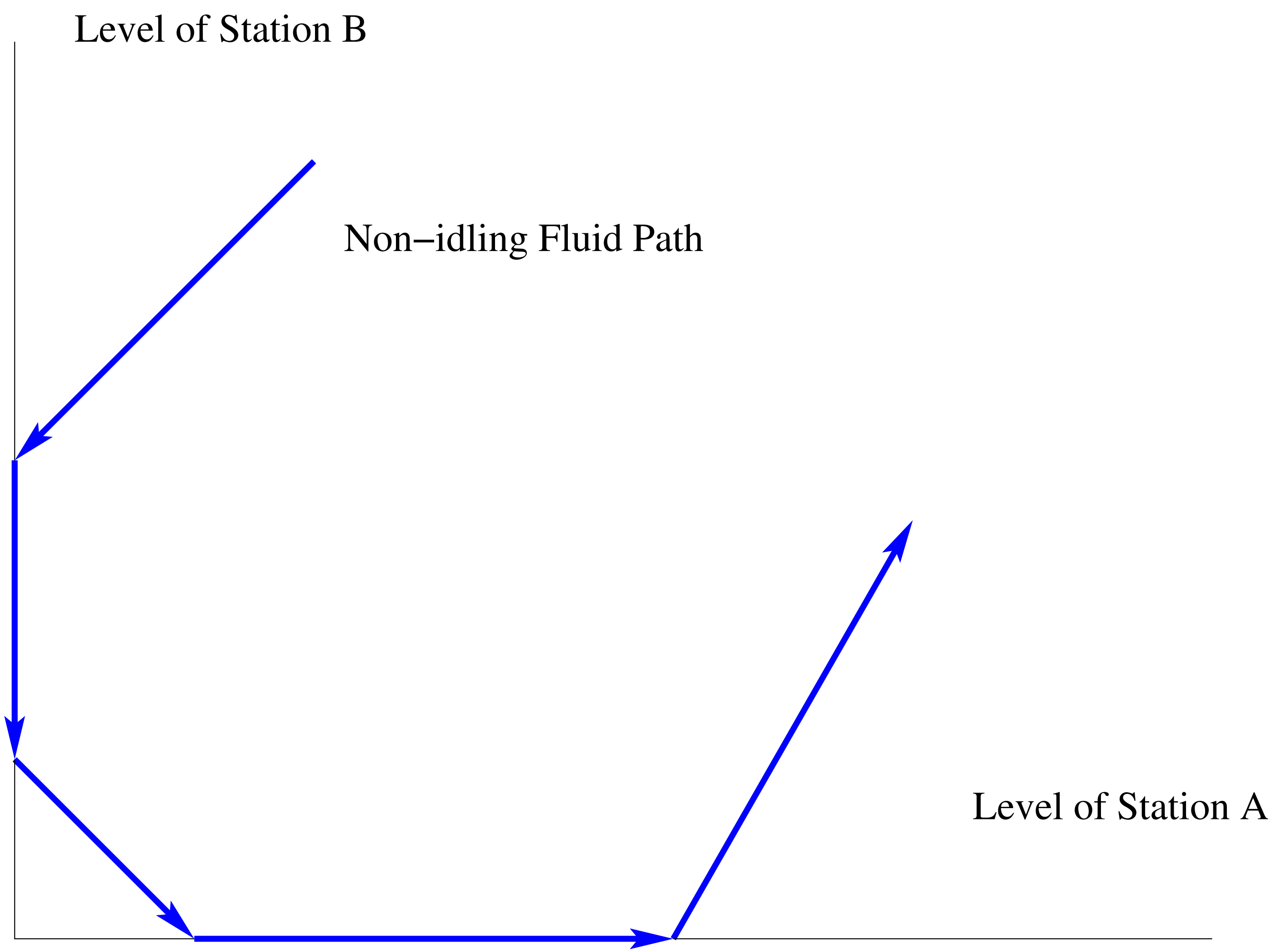
Proof of FDP in 2D



Proof of FDP in 2D



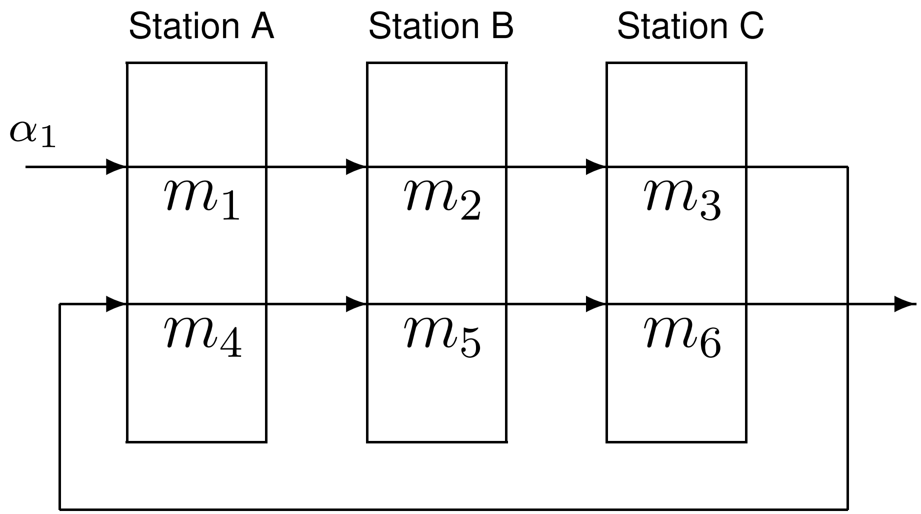
Proof of FDP in 2D



General FDP

- **Open Problem**
  - **Prove or Disprove:** FDP holds for all fluid networks with three or more stations.
  - Can be shown to hold for some three station networks.
  - 3-d geometry ruins the 2-d proof.
  - If **Prove** is possible, then the stability converse holds for networks of arbitrary size.

FDP for Three Stations



Theorem 5: (H and Yildirim 04)

FDP holds for the fluid network above if:

$$m_1 > m_2 > m_3$$

$$m_4 > m_5 > m_6.$$

Proof: Any fluid trajectory can be made non-oscillating.

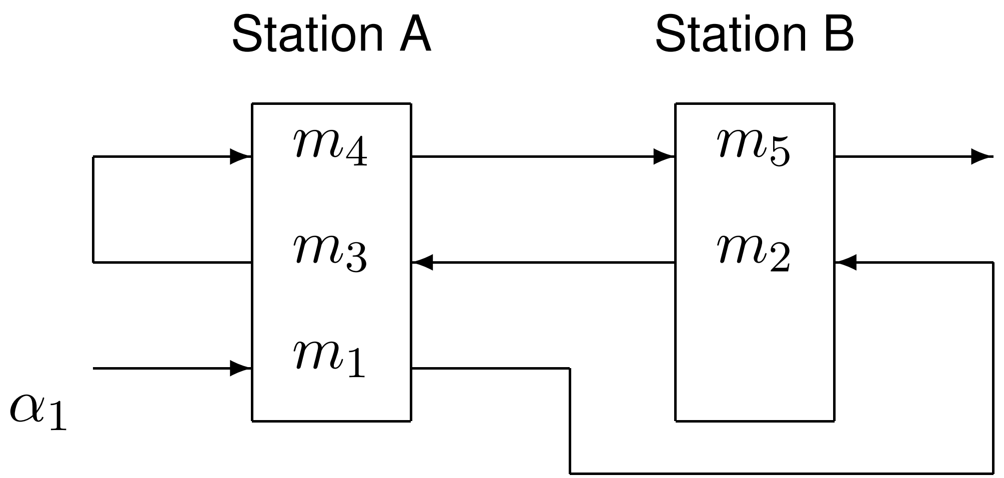
Can we expect a “better” converse?

- **Theorem 3** converse holds for a class of scheduling policies.
- Can the fluid program work for (arbitrary) specific policies?
  - Pick a specific scheduling discipline.
  - Formulate the corresponding fluid model.
  - Use fluid model to determine tight (necessary and sufficient) stochastic stability conditions.
- Dai, H, VandeVate 04 proved that the above program is not possible in general.

Conclusions and Future Work

- Goal: overall view of the limits of fluid stability analysis
- **Open Problems:**
  - Prove or disprove FDP for networks with 3 or more stations $\Rightarrow$ full converse for networks of any size.
  - Fluid network stability still not fully understood
  - Full characterization of “throughput optimal policies”
  - Extend results to Harrison’s stochastic processing networks, telecom models, etc.

A Counterexample



- Consider network with fixed mean value parameters. $\alpha = 1$ and $m = (0.4, 0.1, 0.4, 0.1, 0.4)$.
- Operate under the *static buffer priority* (SBP) discipline: $\{(1, 3, 4), (5, 2)\}$.
- Distributions: exponential, constant.

A Counterexample

- **Theorem 4:** (Dai, H, VandeVate 03) if all distributions are exponential, then $||Q(t)|| \to \infty$ w.p. 1 from any initial state.
- In particular the queueing network is transient.
- **Theorem 5:** if all distributions are constant, then from any initial state, the network eventually enters an orbit. In particular $||Q(t)|| \leq 2$ for all $t \geq T$ for some $T < \infty$.
- The deterministic network is stable in a strong sense.

No General Converse

- **Corollary:** no method (including the fluid model) which takes only mean value as data can sharply determine stability for arbitrary multiclass networks under specific policies.
- No general converse to stability theorems possible!

Proof Outline

- **Theorem 4** – Exponential Case
  - For the counterexample, can show that the fluid model is unstable (there exists a linearly divergent solution).
  - Use large deviations bounds, show that exponential network follows fluid solution, with high probability.
- **Theorem 5** – Deterministic Case
  - Deterministic network is a simple dynamical system.
  - Tedious analysis of trajectories from every initial starting configuration.

Fluid Limits

- $Q_k(t) =$ number of jobs in buffer k at time t.
- $T_k(t) =$ amount of time devoted to processing class k jobs in $[0, t]$.
- Consider SLLN type scaling for the network processes

$$\bar{Q}(t) = \lim_{n\to\infty} \Big(\frac{Q_1(nt)}{n}, \ldots, \frac{Q_N(nt)}{n}\Big)$$

$$\bar{T}(t) = \lim_{n\to\infty} \Big(\frac{T_1(nt)}{n}, \ldots, \frac{T_N(nt)}{n}\Big)$$

- Let $\mathcal{F}(\lim)$ be the set of fluid limits.