Bounds Implied by Drift and Applications

Bruce Hajek

Department of Electrical and Computer Engineering and the Coordinated Science Laboratory

University of Illinois at Urbana-Champaign

Outline

I. Overview of bounds implied by drift

- Law of large numbers and large deviation theory
- Exponential bounds from drift
- Continuous time – comparison of diffusions
- Kingman moment bound and a variation

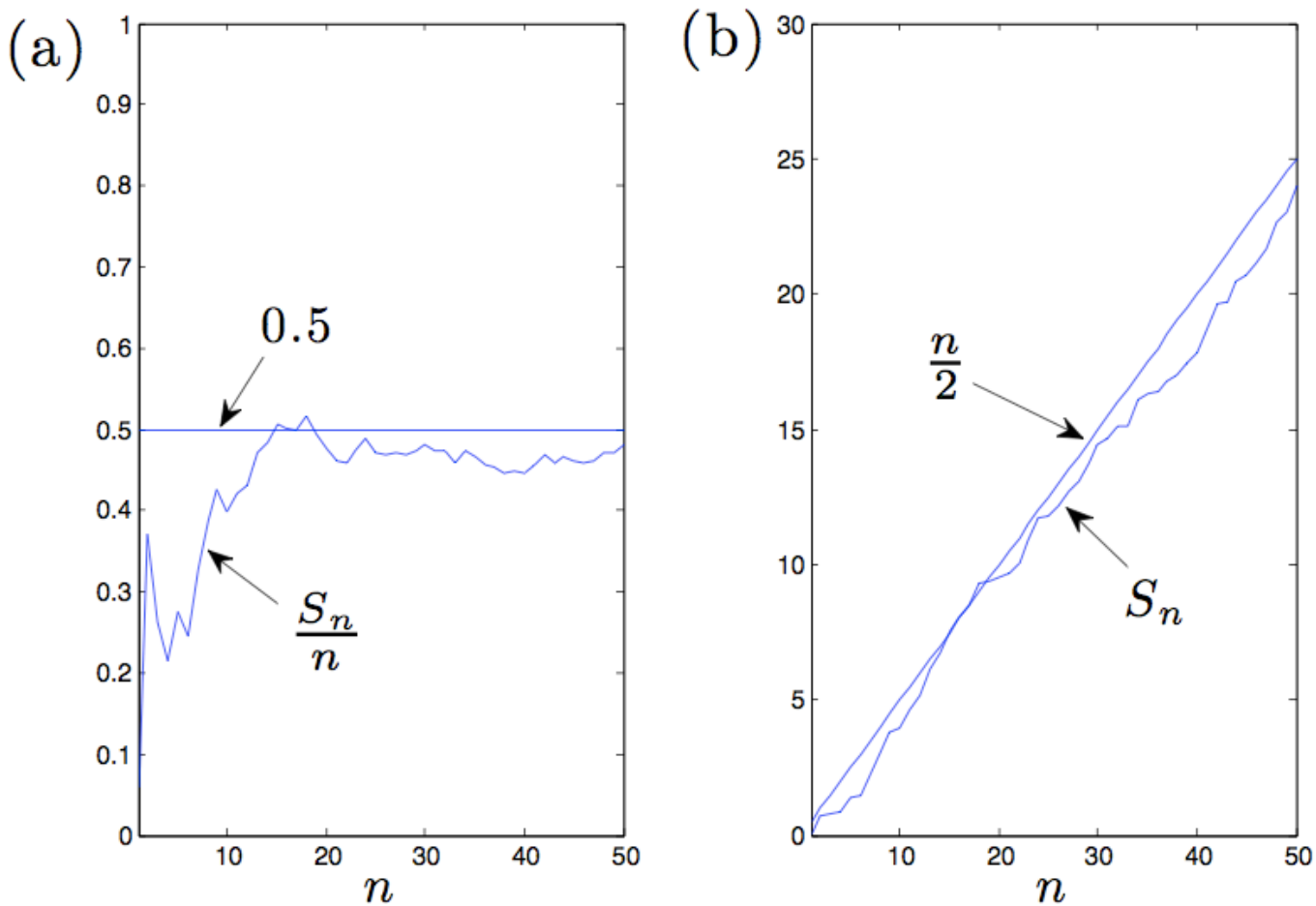
II. Applications

- State space collapse, stability of multiple access protocols and dynamic routing
- Simulated annealing
- Stability of scheduling for constrained traffic

Law of large numbers and large deviations

(some classical implications of drift)

$$S_n = X_1 + \ldots + X_n, \text{where } E[X_i] = \mu.$$



Law of large numbers and large deviations

(some classical implications of drift)

$$S_n = X_1 + \ldots + X_n, \text{where } E[X_i] = \mu.$$

Important consequences:

$$\frac{S_n}{n} \to \mu \quad \text{Law of large numbers}$$

$$P\left\{\frac{S_n}{n} \geq a\right\} \leq e^{-n\ell(a)} \quad \text{large deviation bounds}$$

iid sequences, ergodic sequences (Gartner-Ellis),
sample path ("functional") versions (e.g. Kurt's theorem)

Bounds not depending on exact distributions
(Bernstein inequality, Azuma-Hoeffding inequality, . . .)

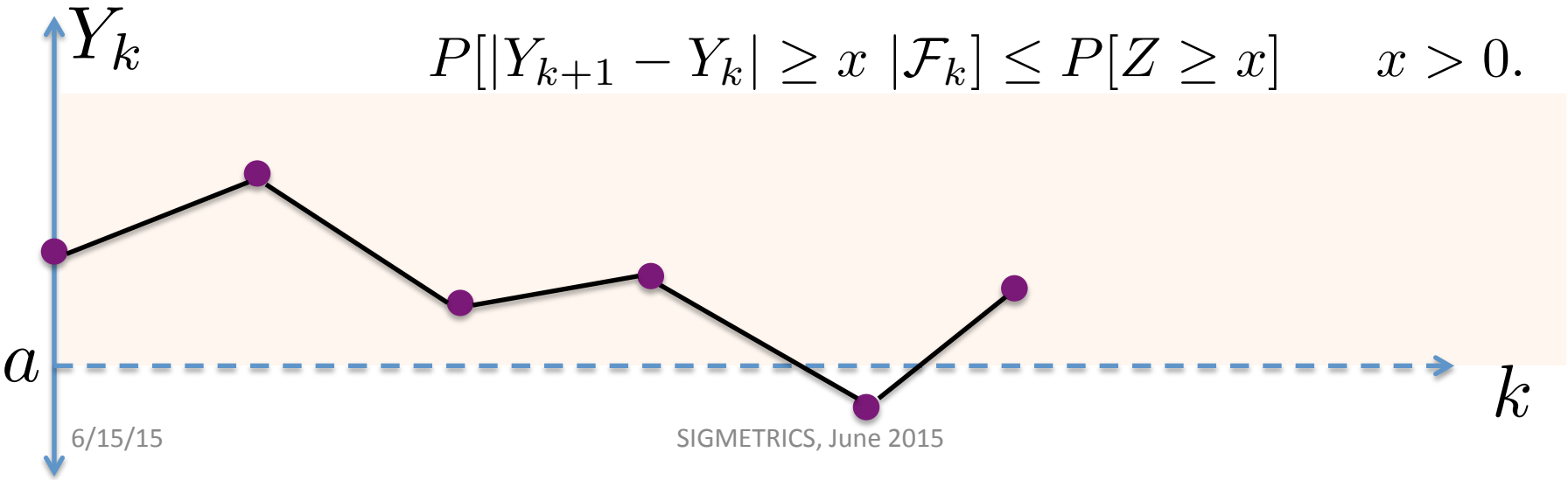
Exponential Bounds from Drift

Some less classical implications of drift

- Let $Y_0, Y_1, \ldots$ be a sequence of random variables.
- $\mathcal{F}_k$ = information at time k
- $E[Y_{k+1} - Y_k | \mathcal{F}_k]$ is the *drift* at instant k

Condition C.1 $E[Y_{k+1} - Y_k | \mathcal{F}_k] \leq -\epsilon_o$ whenever $Y_k \geq a$
Condition C.2 There is a random variable Z with $E[\exp(\lambda Z)] = D$ so that

$$P[|Y_{k+1} - Y_k| \geq x \; |\mathcal{F}_k] \leq P[Z \geq x] \qquad x > 0.$$



Let c, η and ρ be constants such that

$$c \geq \frac{E[\exp(\lambda Z)] - (1 + \lambda E[Z])}{\lambda^2}$$

$$0 < \eta \leq \lambda, \quad \eta < \epsilon_o / c \quad \text{and} \quad \rho = 1 - \epsilon_o \eta + c\eta^2.$$

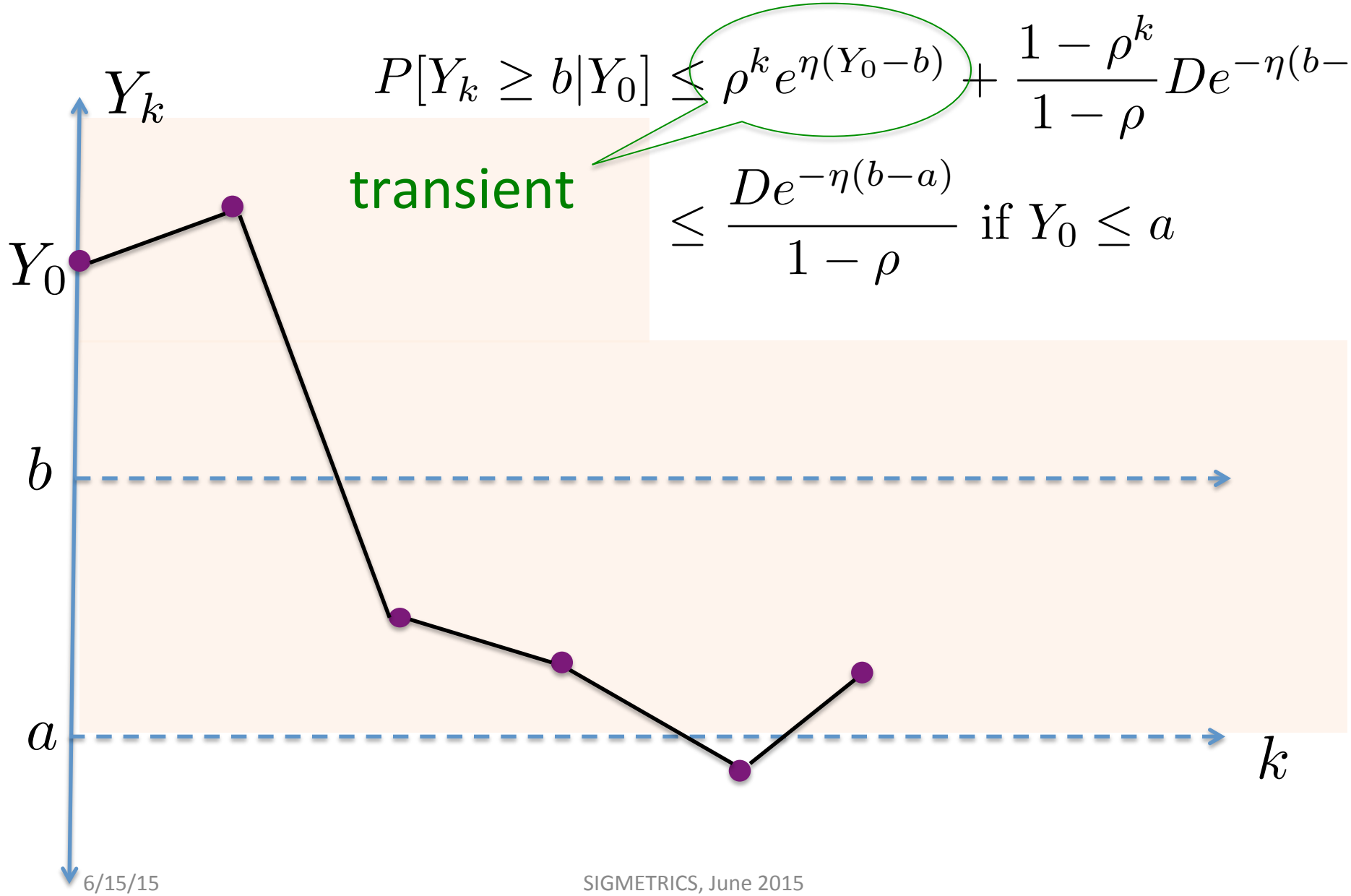
Then $\rho < 1$ and $D \geq 1$

- **Condition D.1** $E[\exp(\eta(Y_{k+1} - Y_k)) I_{\{Y_k \geq a\}} | \mathcal{F}_k] \leq \rho$
- **Condition D.2** $E[\exp(\eta(Y_{k+1} - a)) I_{\{Y_k < a\}} | \mathcal{F}_k] \leq D$

Proposition [H '82] D.1 & D.2 (or C.1 & C.2) imply:

$$P[Y_k \geq b | Y_0] \leq \rho^k e^{\eta(Y_0 - b)} + \frac{1-\rho^k}{1-\rho} D e^{-\eta(b-a)}$$

$$\leq \frac{D e^{-\eta(b-a)}}{1-\rho} \text{ if } Y_0 \leq a$$



Proof

$$
\begin{aligned}
E[e^{\eta Y_{k+1}}|Y_k] &= E[e^{\eta(Y_{k+1}-Y_k)} I_{\{Y_k \geq a\}}|Y_k] e^{\eta Y_k} \\
&\qquad + E[e^{\eta(Y_{k+1}-a)} I_{\{Y_k < a\}}|Y_k] e^{\eta a} \\
&\leq \rho e^{\eta Y_k} + D e^{\eta a},
\end{aligned}
$$

so that

$$E[e^{\eta Y_{k+1}}|Y_0] \leq \rho E[e^{\eta Y_k}|Y_0] + D e^{\eta a}.$$

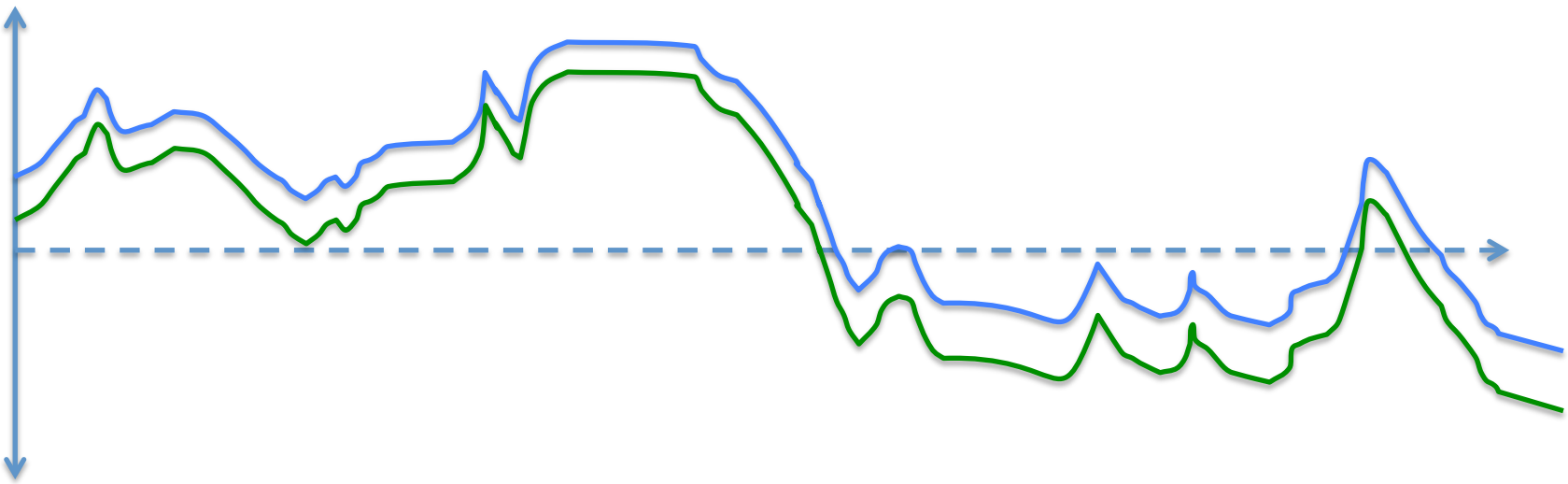
Then by induction on k :

$$E[e^{\eta Y_k}|Y_0] \leq \rho^k e^{\eta Y_0} + \frac{1-\rho^k}{1-\rho} D e^{\eta a}.$$

Chernoff bound: $P[Y_k \geq b|Y_0] \leq e^{-\eta b} E[e^{\eta Y_k}|Y_0]$ completes proof.

Continuous time

comparison of diffusions



Theorem (Classical comparison theorems, see Ikeda & Watanabe)
Suppose σ, b_1 and b_2 are Lipschitz continuous. For $i = 1, 2$, let

$$dX_t^i = \sigma(X_t^i)dw_t + b_i(X_t^i)dt$$

If $b_1 \geq b_2$ and $X_0^1 \geq X_0^2$ then $X_t^1 \geq X_t^2$ for all t.
Proof idea: Construct processes jointly and invoke strong uniqueness.

Continuous time

comparison of diffusions (contd.)

Theorem ([H, '85]) Let x and y be continuous semimartingales:

$$\begin{aligned} dx_t &= \mu_t(X_t)dt + \sigma_t dw_t \\ dy_t &= m dt + \rho dv_t \end{aligned}$$

Suppose: $\mu_t \leq m$, $|\sigma_t| \leq \rho$, and $x_0 \leq y_0$.
Then: $P\{x_t \geq c\} \leq 2P\{y_t \geq c\}$ $\forall c$, and $E[\Phi(x_t)] \leq E[\Phi(y_t)]$
for any nondecreasing, convex functions Φ.
Proof idea: $x_t \leq z_t \stackrel{\triangle}{=} \frac{Y_t^1 + Y_t^2}{2}$ and $y_t \stackrel{d.}{=} Y_t^1 \stackrel{d.}{=} Y_t^2$, where

$$Y_t^i = y_0 + mt + \left[\int_0^t \sigma_s dw_s + (-1)^i \int_0^t (\rho^2 - \sigma_s^2)^{1/2} d\tilde{w}_s \right].$$

Extensions: reflections, diffusion coefficient depending on state

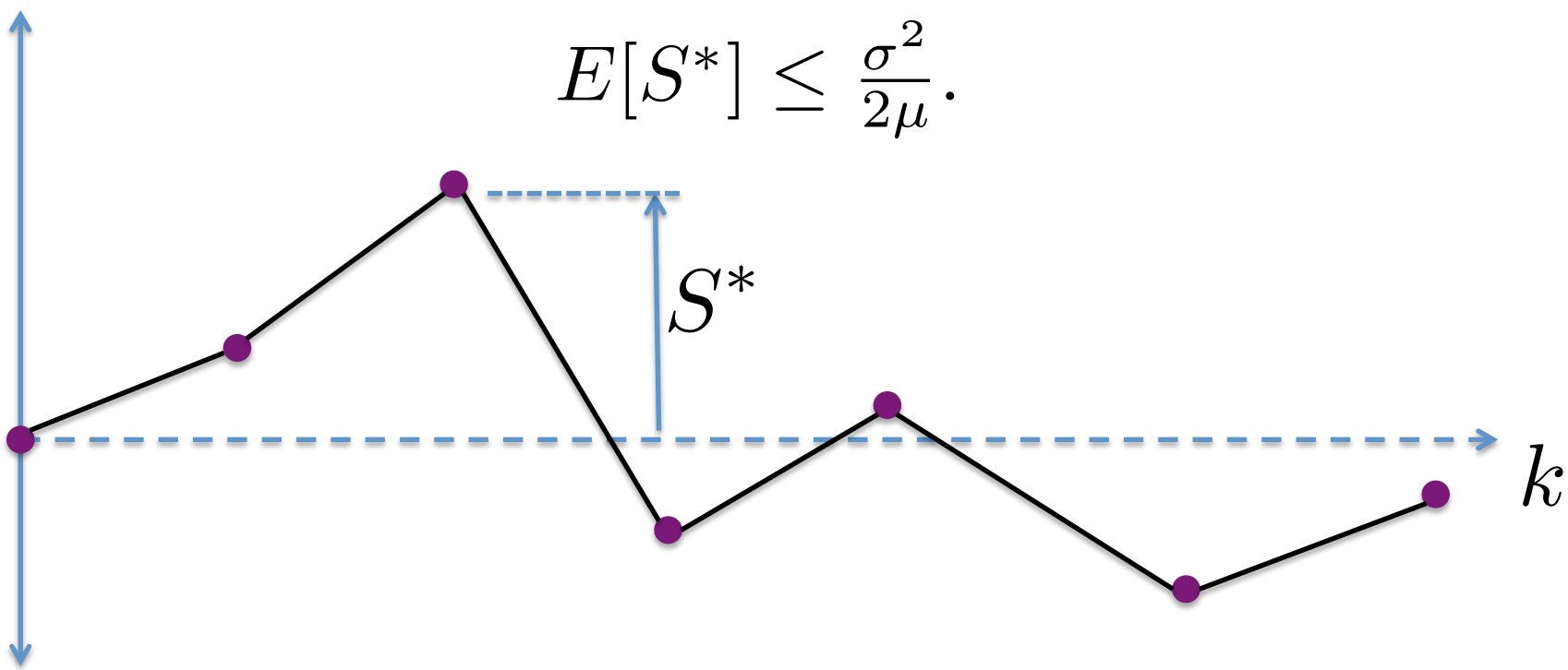
Kingman's moment bound for random walks,

and a variation

Let $S_0 = 0$ and $S_k = U_1 + \cdots + U_k$ where $U_1, U_2, \ldots$ are iid with mean $-\mu$ and variance $\sigma^2 < \infty$.

Let $S^* = \sup_{k \geq 0} S_k$. Kingman's bound ['62] is

$$E[S^*] \leq \frac{\sigma^2}{2\mu}.$$



Proof of Kingman's bound

Let $W_n = \max\{0, U_1, U_1 + U_2, \cdots, U_1 + \cdots + U_n\}$.
Then $W_n \nearrow S^*$ as $n \nearrow \infty$, so $E[W_n] \nearrow E[S^*]$.

$$W_{n+1} = \max\{0, U_1 + \underbrace{\max\{0, U_2, U_2 + U_3, U_2 + \cdots + U_{n+1}\}}_{\text{same distribution as } W_n}\}.$$

so

$$W_{n+1} \stackrel{d.}{=} (U + W_n)_+.$$

Trivially,

$$U + W_n = (U + W_n)_+ - (U + W_n)_-$$

Take expectations, use $E[W_{n+1}] \geq E[W_n] : E[(U + W_n)_-] \geq \mu$.
Square, use $(a)_+(a)_- \equiv 0$, taking expectations,
use $E[W_{n+1}^2] \geq E[W_n^2]$: $E[U^2] - 2\mu E[W_n] \geq E[(U + W_n)_-^2]$.

Proof of Kingman's bound (continued)

Yields

$$
\begin{aligned}
E[W_n] &\leq \frac{E[U^2] - E[(U+W_n)_-^2]}{2\mu} \\
&\leq \frac{E[U^2] - \mu^2 - E[(U+W_n)_-^2] + E[(U+W_n)_-]^2}{2\mu} \\
&= \frac{\sigma^2 - \mathrm{Var}((U+W_n)_-)}{2\mu} \leq \frac{\sigma^2}{2\mu}
\end{aligned}
$$

Finally, letting $n \to \infty$ yields Kingman's bound.

Return to Kingman's bound:

Let $S_0 = 0$ and $S_k = U_1 + \cdots + U_k$ where $U_1, U_2, \ldots$ are iid with mean $-\mu$ and variance $\sigma^2 < \infty$.
Let $S^* = \sup_{k \geq 0} S_k$. Kingman's bound ['62] is

$$E[S^*] \leq \frac{\sigma^2}{2\mu}.$$

Q. What if we drop assumption of identical distribution?

Kingman's bound in continuous time (proof is by simple limit procedure)

Let Y be a stationary independent increment (SII) process with $Y_0 = 0$ such that

$$E[Y_t] = -\mu t \quad \text{and} \quad \mathrm{Var}(Y_t) = \sigma^2 t.$$

Then $E[Y^*] \leq \frac{\sigma^2}{2\mu}$.

By Markov inequality, $P\{Y^* \geq a\} \leq \frac{1}{\gamma a}$ where $\gamma = \frac{\mu}{\sigma^2}$.

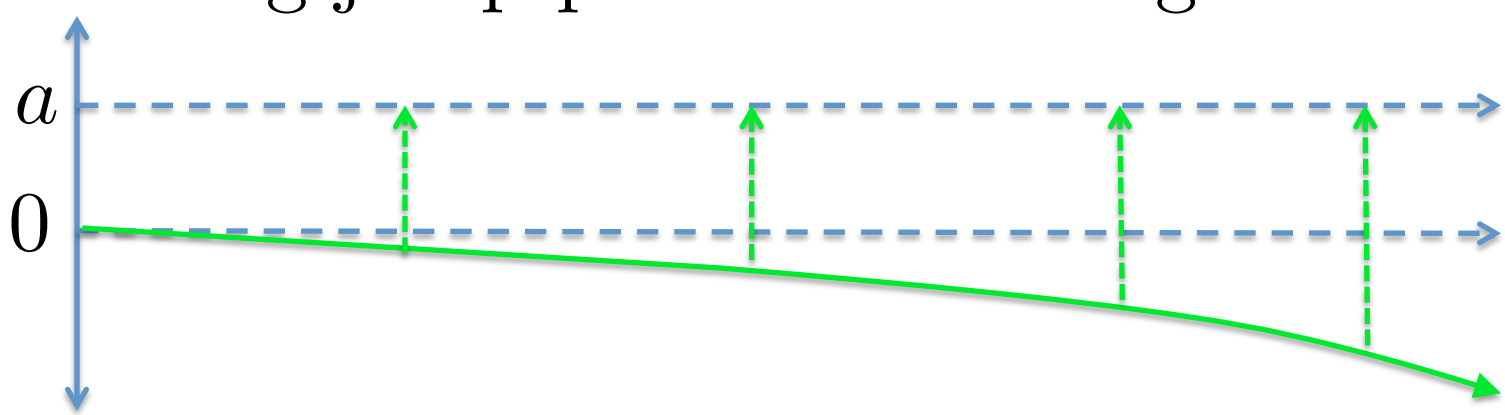
Q. What if we drop assumption of identical distribution?

Dropping identical distribution assumption from Kingman's bound

Proposition 1 [H, '14] Suppose $Y_0 = 0$, and both $(Y_t + \mu t : t \geq 0)$ and $(Y_t^2 - \sigma^2 t : t \geq 0)$ are supermartingales (i.e. have negative drift).

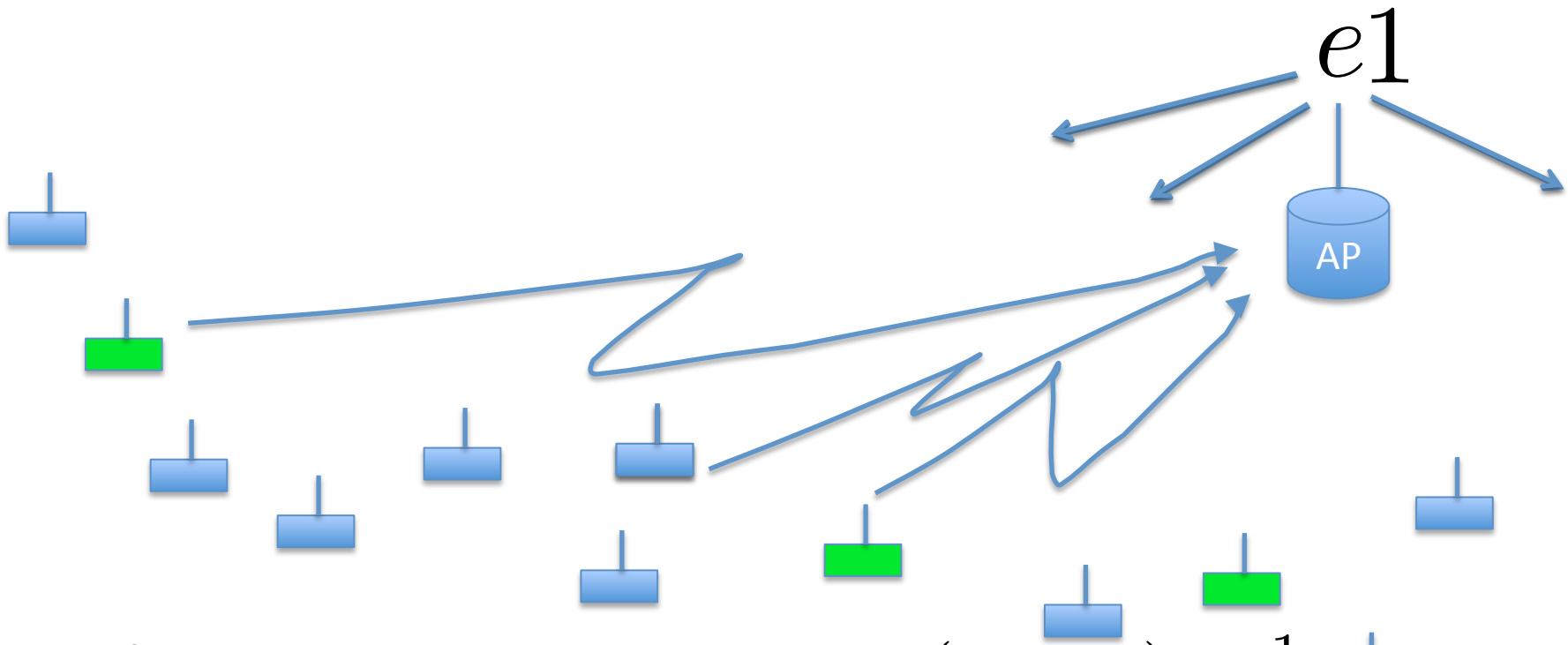
(a) For $a \geq 0$: $\quad P\{Y^* \geq a\} \leq \frac{1}{1+\gamma a}$ where $\gamma = \frac{\mu}{\sigma^2}$.

(b) Equality holds if and only if Y is the big jump process with target a.



Application: State space collapse and stability analysis

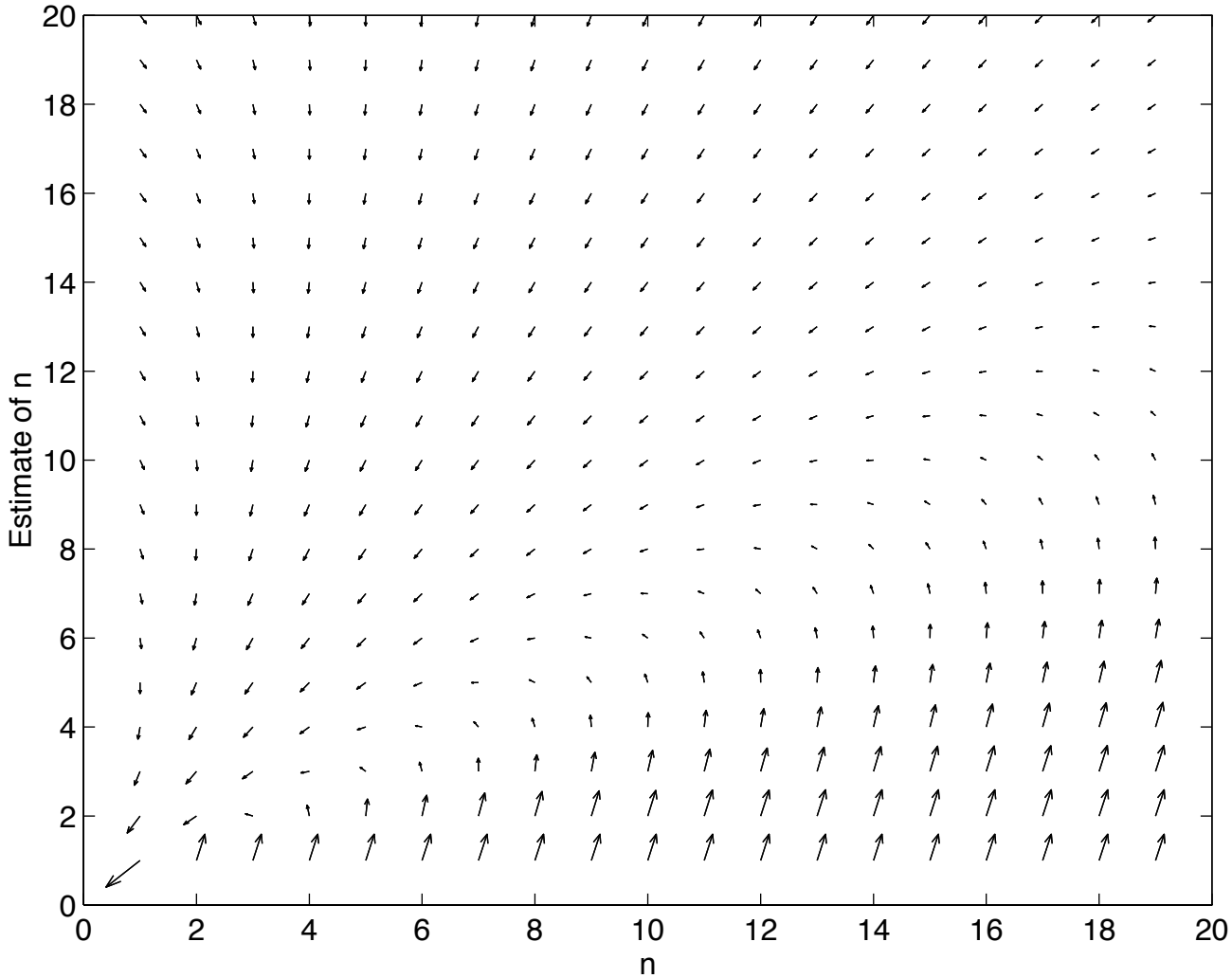
Stabilizing ALOHA ([H & van Loon '82],[Mikhailov '79])



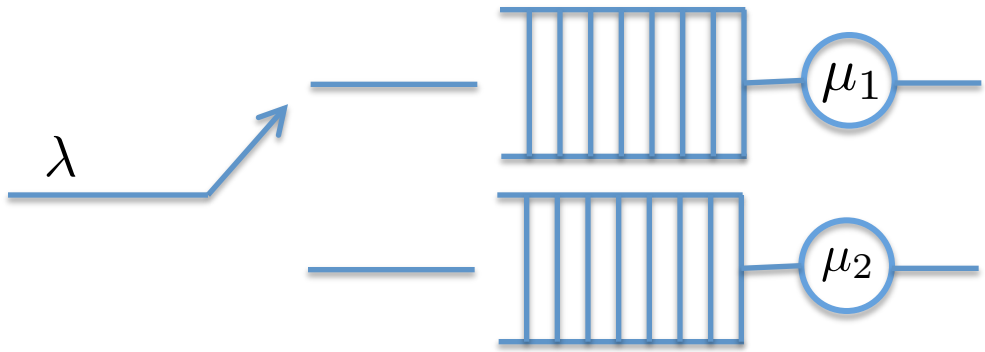
0-1-e feedback; $P_{success} = p(1-p)^{n-1}$.

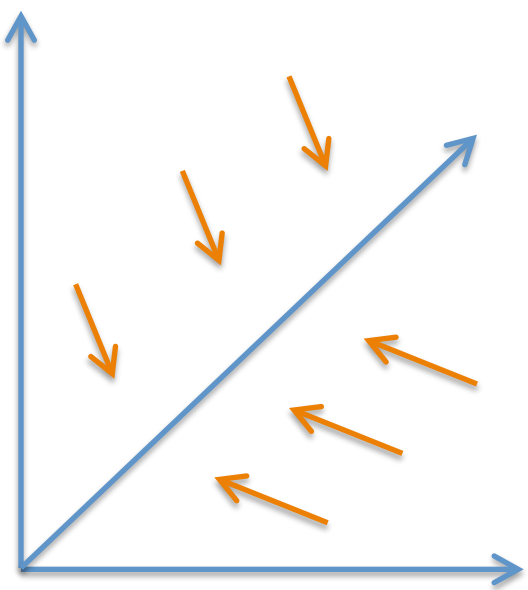
Use $p = \frac{1}{\hat{n}}$.

Joint drift: backlog and estimated backlog for multiple access



State space collapse for join the shortest queue

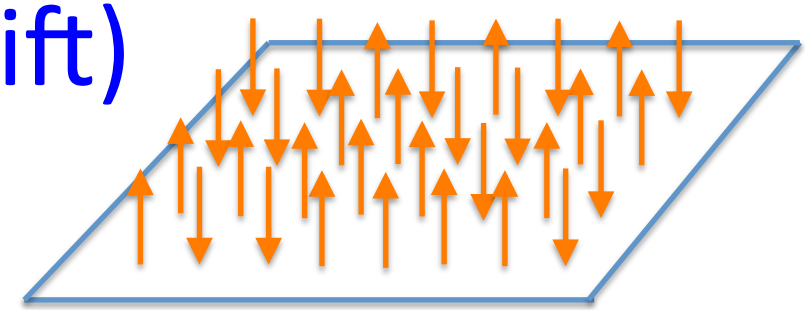




$$\text{drift}(N_1 - N_2) = \mu_2 - \mu_1 \pm \lambda$$

Eryilmaz Srikant 2012

Simulated Annealing (insights from drift)



$$V(s) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j - \sum_i h_i s_i \quad \text{(Ising model)}$$

$$\pi(s) = \frac{1}{Z} \exp\left(-\frac{1}{T} V(s)\right)$$

Recent work of Jiang and Walrand using MCMC for CSMA is related to simulated annealing

Goal: Find global minimizer of *V* over all possible states s

Simulated Annealing Algorithm:

Initialize state: $s \leftarrow s_0$

For $k = 1$ to $k_{\max}$:

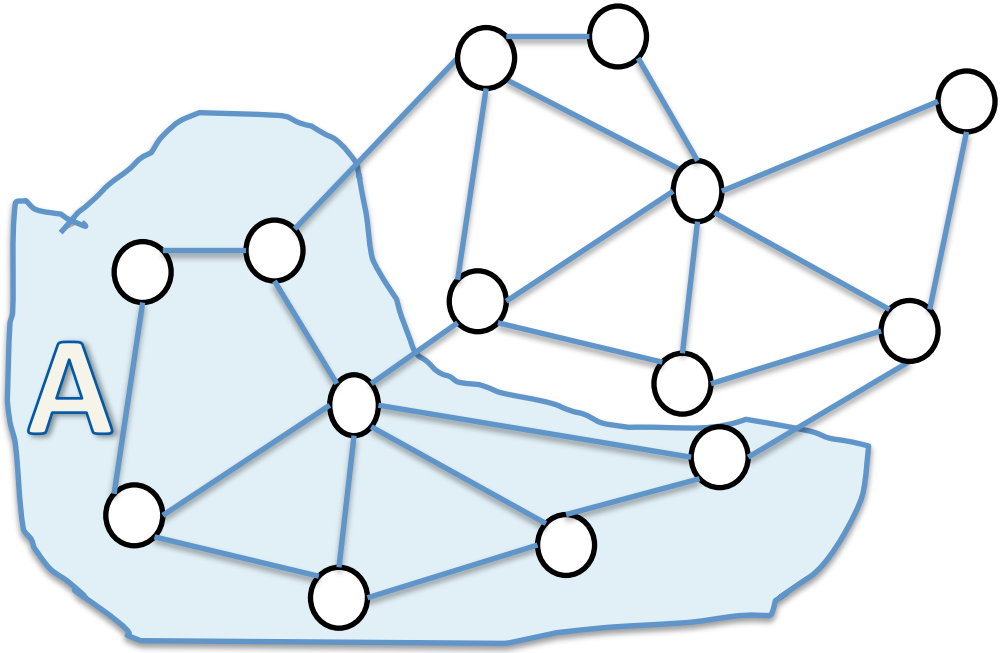
Choose a neighbor s' of s

Compute $\triangle V = V(s') - V(s)$ (change in cost)

If $\triangle V \leq 0$, $s \leftarrow s'$

If $\triangle V > 0$, $s \leftarrow s'$ with probability $e^{-\triangle V/T_k}$

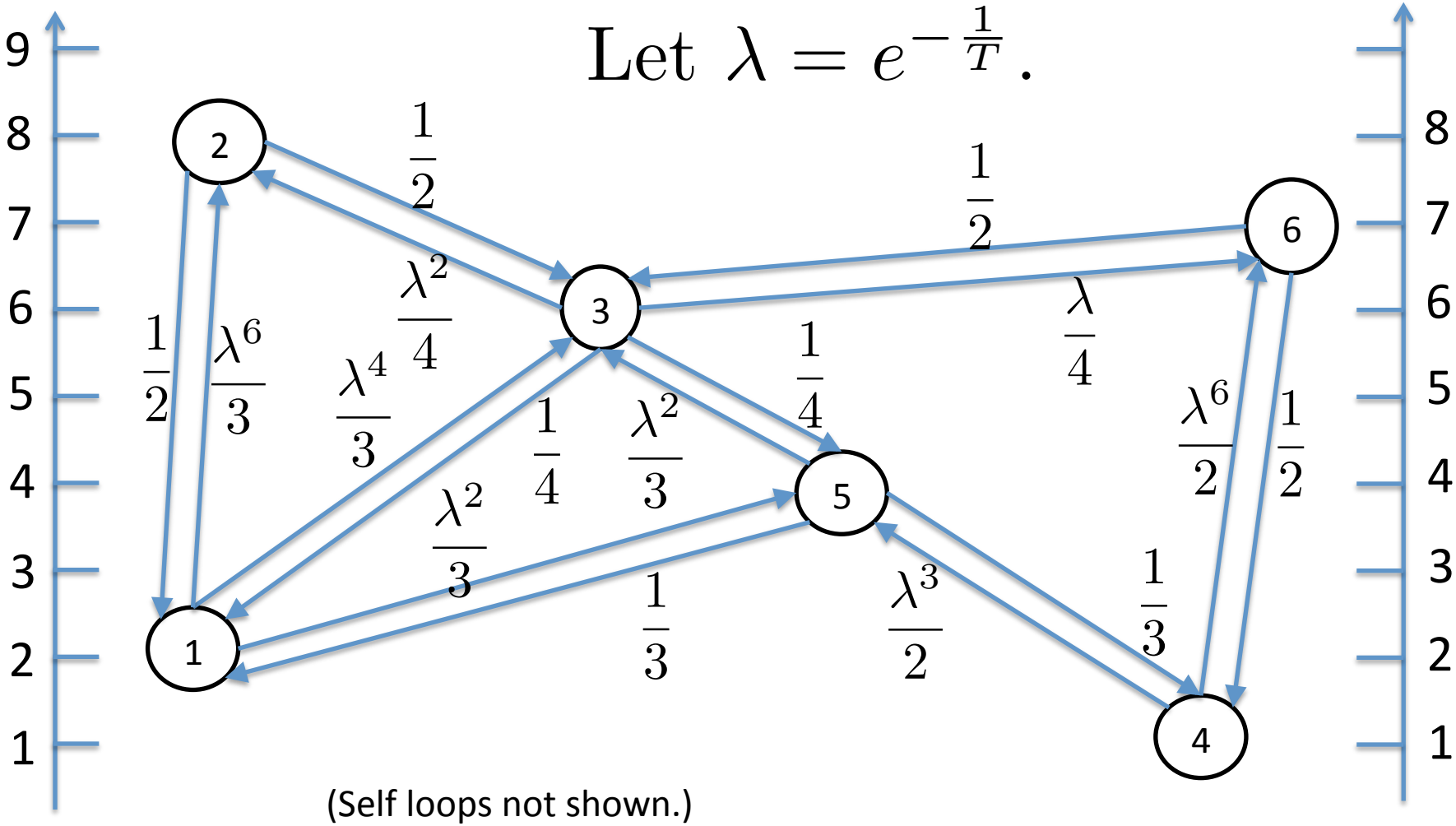
Example: set partitioning



$$V(A, A^c) = e(A, A^c) + \gamma(|A| - |A|^c)^2$$

Markov chain induced by simulated annealing

$$\text{Let } \lambda = e^{-\frac{1}{T}}.$$

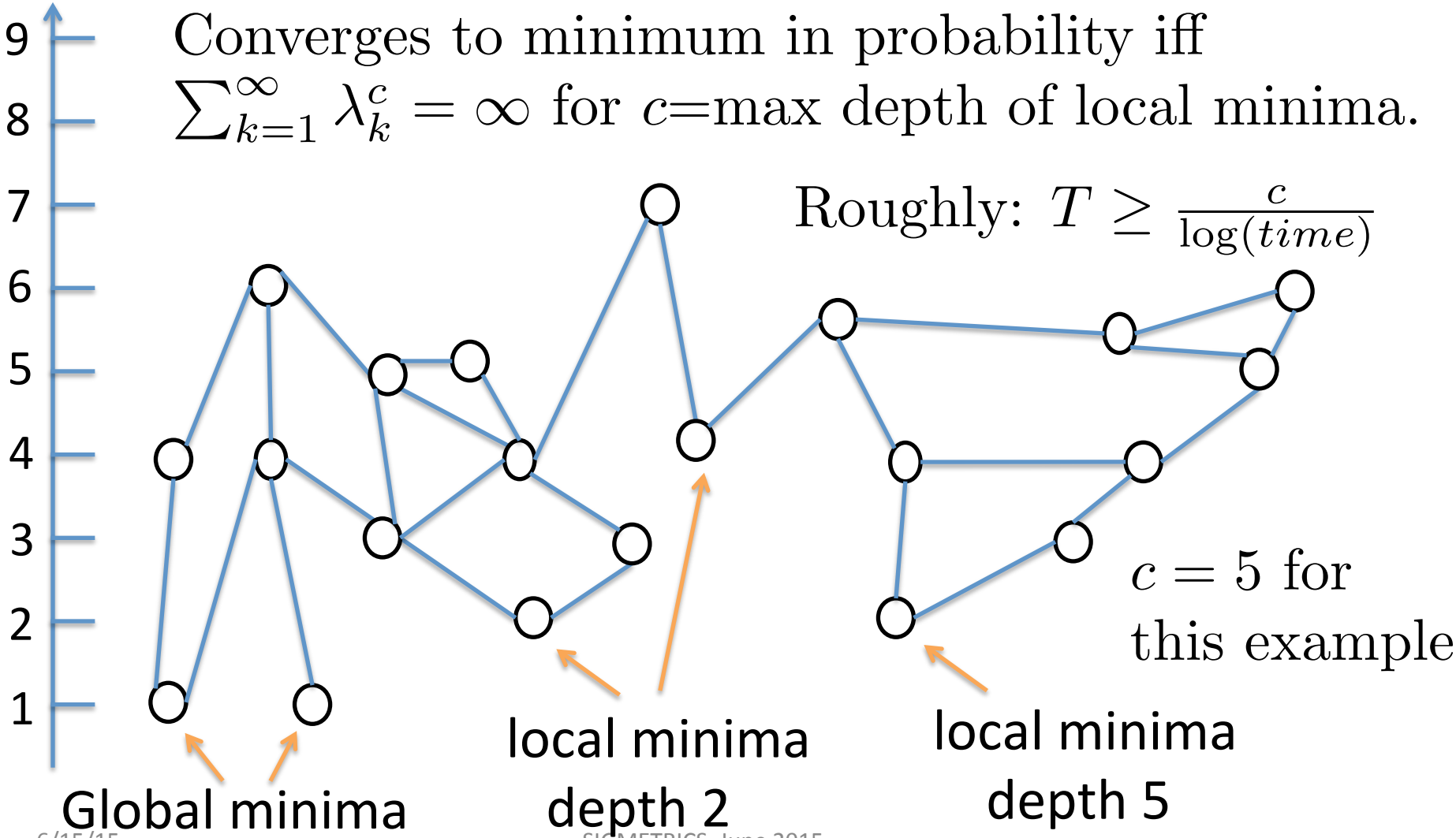


(Self loops not shown.)

Large **time** behavior of simulated annealing

Converges to minimum in probability iff $\sum_{k=1}^{\infty} \lambda_k^c = \infty$ for c=max depth of local minima.

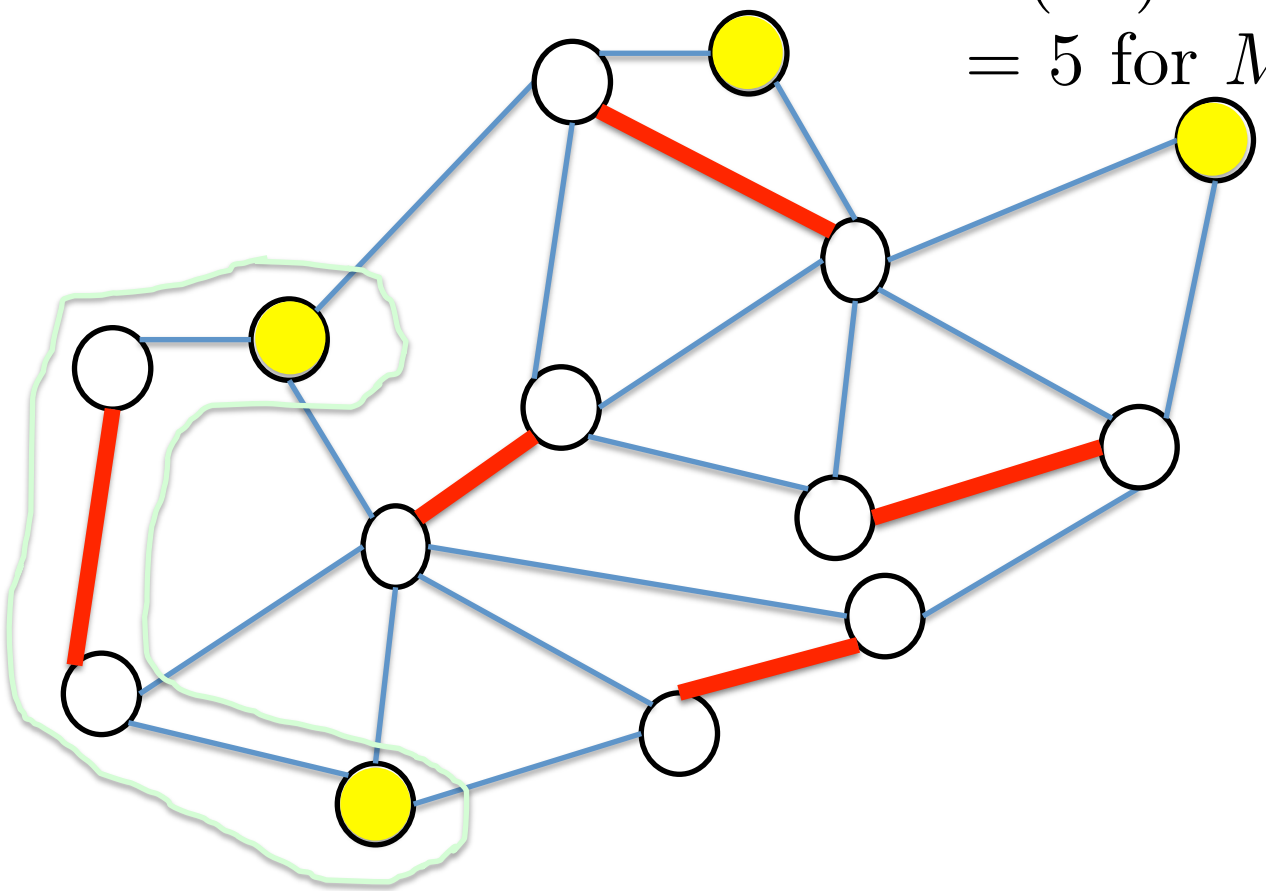
Roughly: $T \geq \frac{c}{\log(time)}$



Large **problem size** behavior of simulated annealing

Addressed in [Sasaki – H. ‘88] for maximum matching problem.

$$V(M) = |M|$$
$$= 5 \text{ for } M \text{ shown.}$$



$$c = 1$$

Large **problem size**
(a positive result for approximate maximum matching)

$d^* =$ maximum vertex degree
$m^* =$ maximum cardinality over all matchings

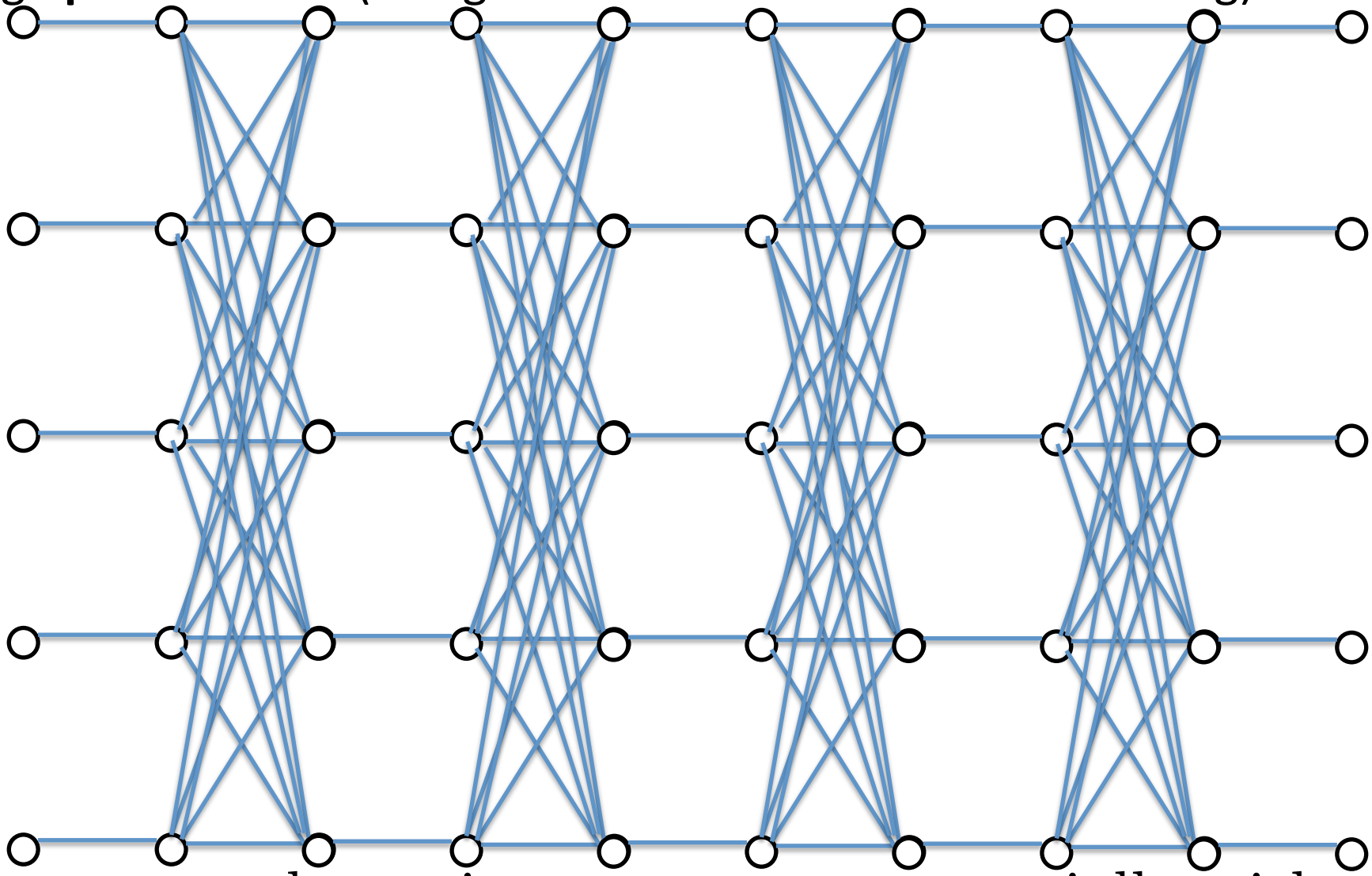
Theorem (Sasaki & H, '88) Simulated annealing at constant temperature

$$T = 1/\log\left(3\left(1+\frac{1}{\epsilon}\right)|V|^2(2d^*)^{1/\epsilon}\right)$$

reaches fraction at least $(1-\epsilon)$ of maximum matching in expected time

$$E[R_\epsilon] \leq 24\left(1+\frac{1}{\epsilon}\right)^2|V|^5(2d^*)^{1/\epsilon}$$

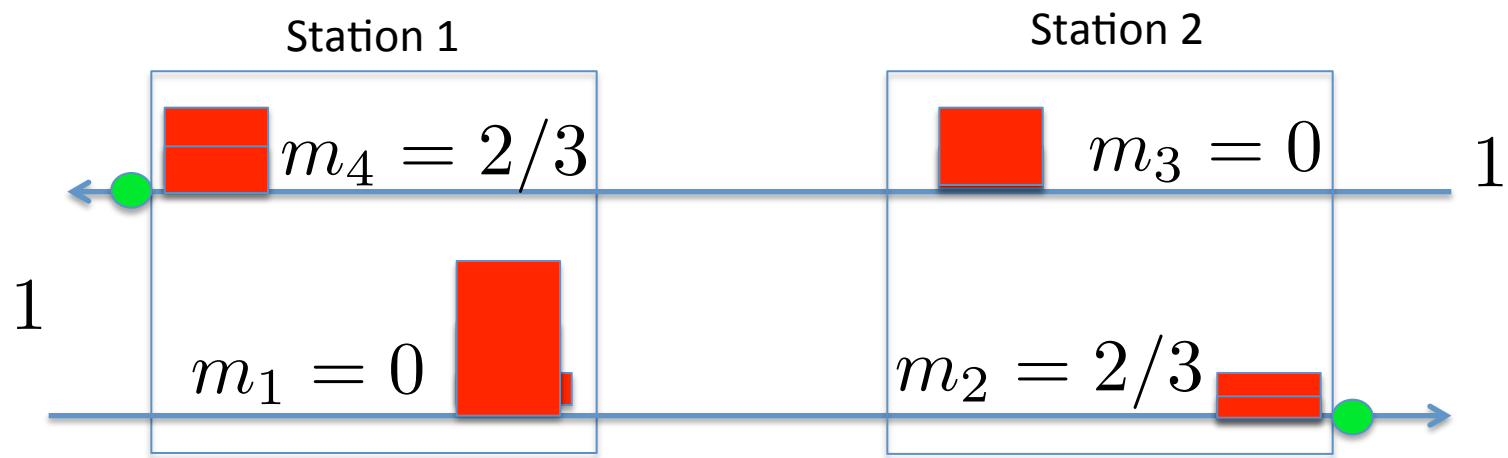
Large **problem size** (a negative result for maximum matching)



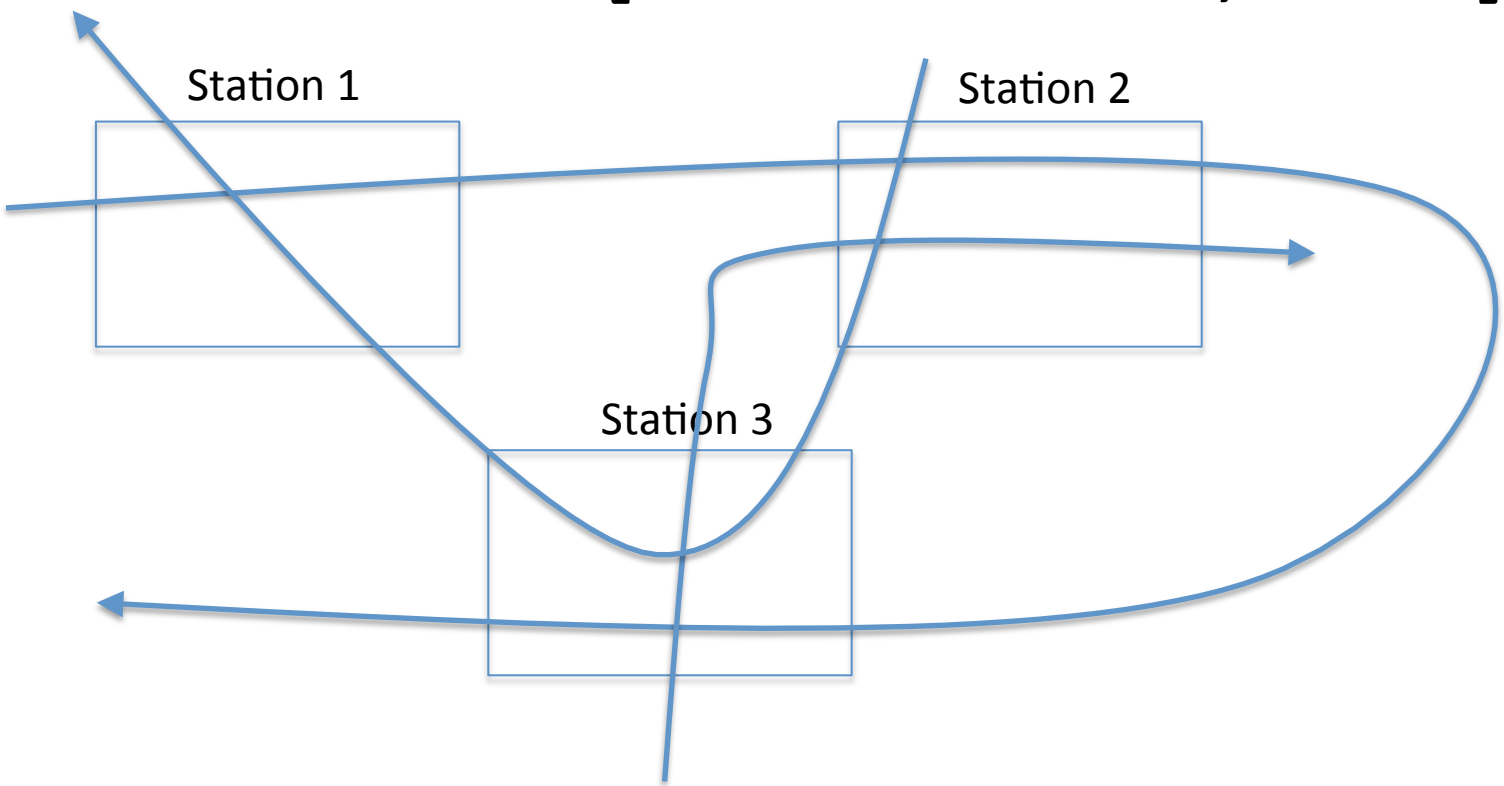
Time to reach maximum grows exponentially with n for any causal control policy for temperature.

Fluid Analysis and Deterministic Constraints (insights from drift)

Rybko & Stolyar example (`92)
(Also see Lu & Kumar example `91)



If system is stable for fluid traffic it is stable for stochastic traffic [Rybko & Stolyar `92, Dai `95] and for deterministically constrained traffic [Gamarnik `00, H`00]



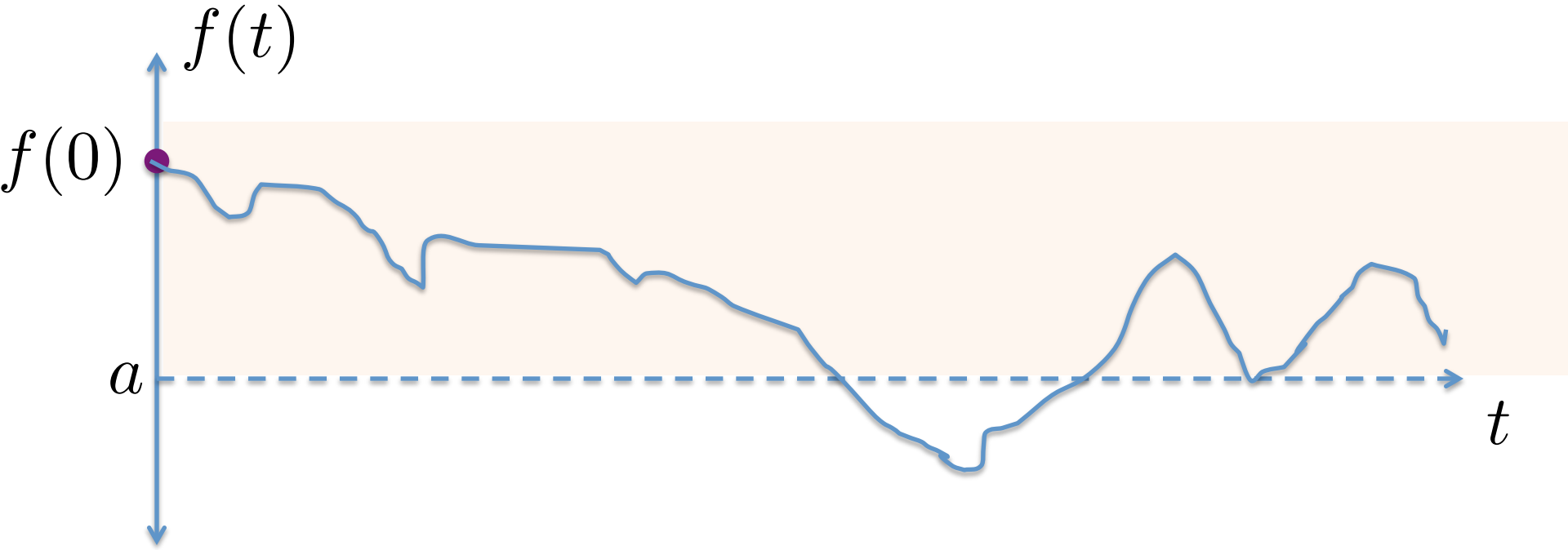
Bound implied by drift for deterministically constrained traffic

Lemma

Let $f : [0, \infty) \to [0, \infty)$ be right-continuous so for some τ, δ, a, b, c, with $\delta < 1$

$$f(t + \tau f(t)) \leq \delta f(t) \quad \text{if } f(t) > a$$
$$f(t + s) - f(s) \leq b + ct \quad \text{if } s, t \geq 0$$

Then $f(s) \leq (a + b)(1 + c\tau/(1 - \delta))$ for $s \geq f(0)\frac{\tau}{1-\delta}$.



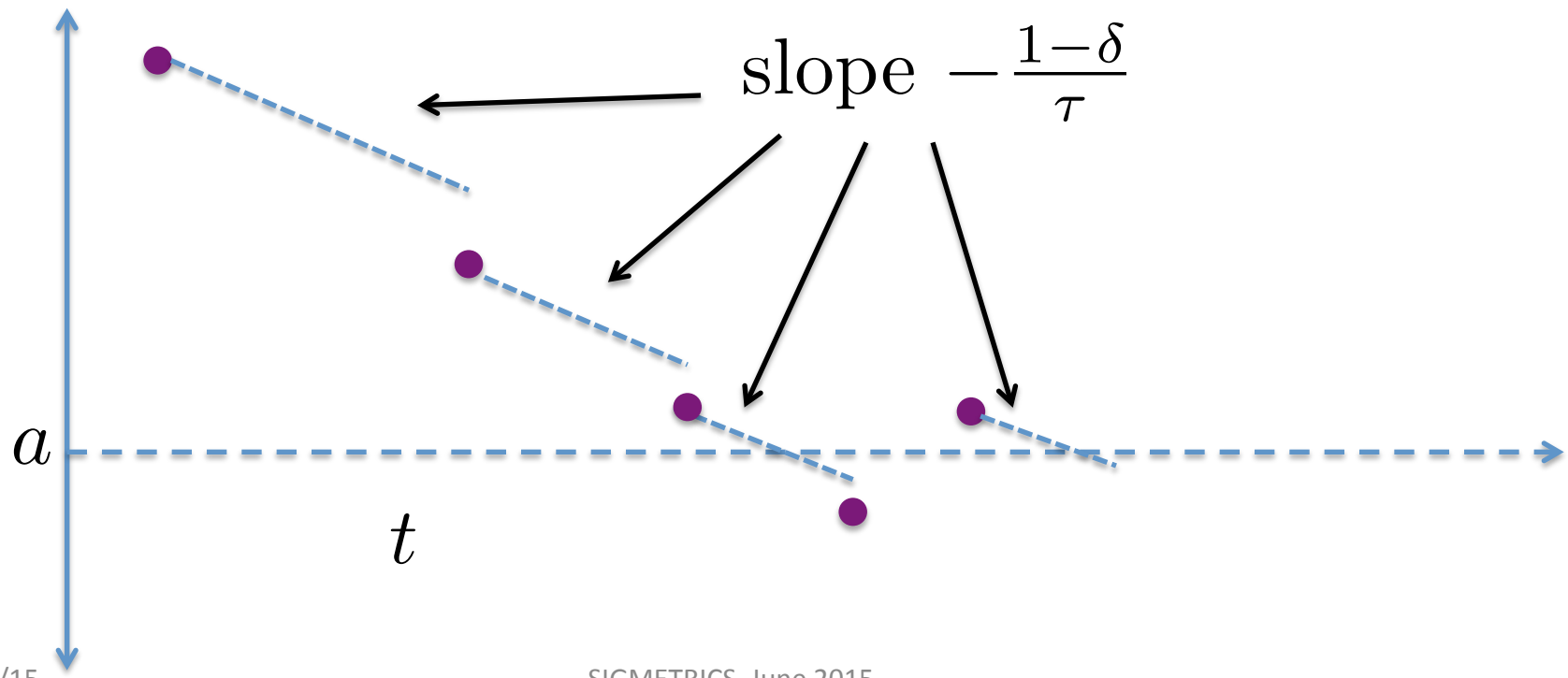
Lemma

Let $f : [0, \infty) \to [0, \infty)$ be right-continuous so for some τ, δ, a, b, c, with $\delta < 1$:

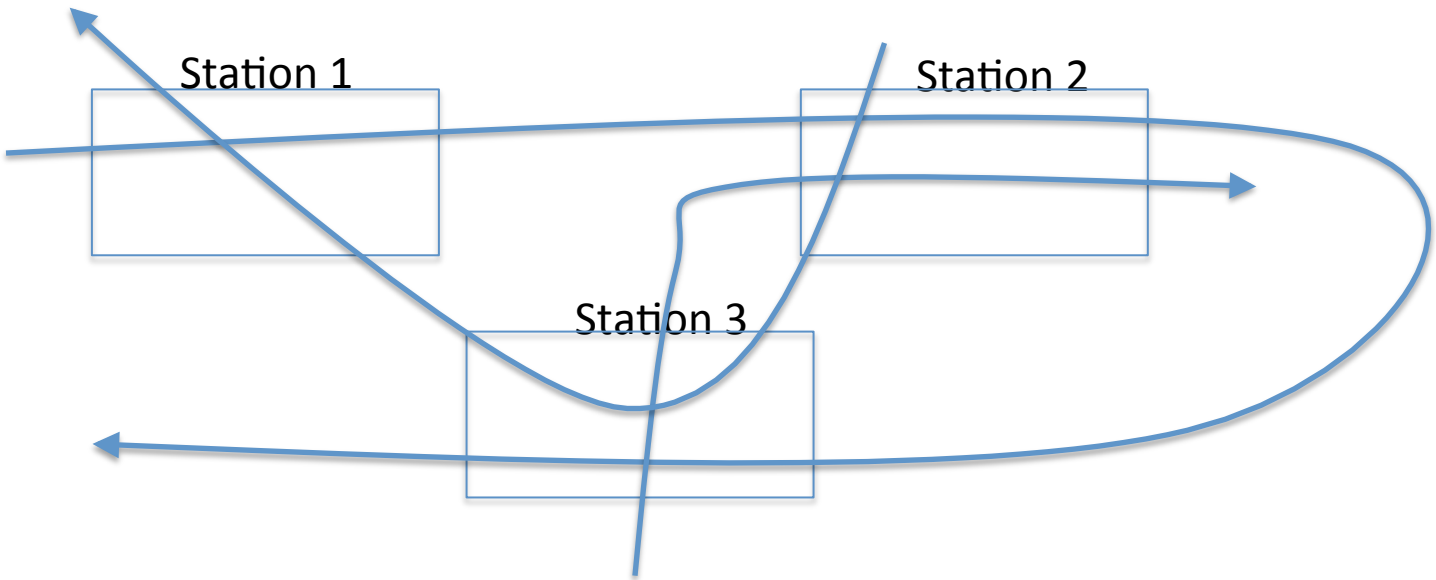
$$f(t + \tau f(t)) \leq \delta f(t) \quad \text{if } f(t) > a$$
$$f(t + s) - f(s) \leq b + ct \quad \text{if } s, t \geq 0$$

Then $f(s) \leq (a + b)(1 + c\tau/(1 - \delta))$ for $s \geq f(0)\frac{\tau}{1-\delta}$.

Idea: $f(t) > a$ implies $f(t + \tau f(t)) \leq f(t) - \frac{1-\delta}{\tau}(\tau f(t))$.



Q: Is it enough to have enough capacity at each station for this network topology?



A: No, because it generalizes Rybko-Stolyar.

But Dai & Weiss `96 has elegant proof of stability for fluid models if all traffic in ring is in same direction. Recovers result of Tassiulas & Georghiades `95.

References

[1] J.G. Dai. On positive harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *The Annals of Applied Probability*, pages 49–77, 1995.

[2] J.G. Dai and G. Weiss. Stability and instability of fluid models for reentrant lines. *Mathematics of Operations Research*, 21(1):115–134, 1996.

[3] A. Eryilmaz and R. Srikant. Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems*, 72(3-4):311–359, 2012.

[4] S. Foss, B. Hajek, and A. Turlikov. Stability of a random multiple access channel with "success-failure" feedback. Available on ArXiv 1304.7626, April 2013.

[5] D. Gamarnik. Using fluid models to prove stability of adversarial queueing networks. *IEEE Trans Automatic Control*, 45(4):741–746, 2000.

[6] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied probability*, 14:502–525, 1982.

[7] B. Hajek. Mean stochastic comparison of diffusions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 68(3):315–329, 1985.

[8] B. Hajek. Cooling schedules for optimal annealing. *Math Operations Research*, 13(2):311–329, 1988.

[9] B. Hajek. Large bursts do not cause instability. *IEEE Trans Automatic Control*, 45(1):116–118, 2000.

[10] B. Hajek. A maximal inequality for supermartingales. *Elect. Commun. Probability*, 19:1–10, 2014.

[11] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes.* Elsevier, 2014.

[12] L. Jiang and J. Walrand. Approaching throughput-optimality in distributed csma scheduling algorithms with collisions. *IEEE/ACM Transactions on Networking*, 19(3):816–829, 2011.

[13] J.F.C. Kingman. Some inequalities for the queue GI/G/1. *Biometrika*, 49(3/4):315–324, December 1962.

[14] J.F.C. Kingman. A martingale inequality in the theory of queues. *Cambridge Phiolos. Soc.*, 59:359–361, 1964.

[15] S.H. Lu and P.R. Kumar. Distributed scheduling based on due dates and buffer priorities. *IEEE Transactions on Automatic Control*, 36(12):1406–1416, 1991.

[16] A.N. Rybko and A.L. Stolyar. Ergodicity of stochastic processes describing the operation of open queueing networks. *Problemy Peredachi Informatsii*, 28(3):3–26, 1992.

[17] G.H. Sasaki and B. Hajek. The time complexity of maximum matching by simulated annealing. *Journal of the ACM (JACM)*, 35(2):387–403, 1988.

[18] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans Automatic Control*, 37(12):1936–1948, 1992.

[19] L. Tassiulas and L. Georgiadis. Any work-conserving policy stabilizes the ring with spatial re-use. *IEEE/ACM Transactions on Networking*, 4(2):205–208, 1996.

Thanks!