

Transform Method for Markov-Modulated Queues

Daniela Hurtado-Lange (Kellogg) and Izzy Groszof (IEMS)
Northwestern University

SNAPP Seminar — December 1st, 2025

Waiting in Line



C'mon, do something

Ultimate goal: Minimize Delay

SWIFT
TOUR

FRONTIER
TAYLOR SWIFT
TOURING

GENERAL PUBLIC ON SALE

CONCERTS STARTS FRI 30 JUN, 2PM (AEST)

Your turn to purchase tickets is coming soon

Next update in 4 seconds

Line Sitter and Waitlist Services in

★★★★★

Queuing up for...
energy. Hire s...

✓ Yes, line wa...
world!

✓ Your profession...
as they wait in li...
ated.

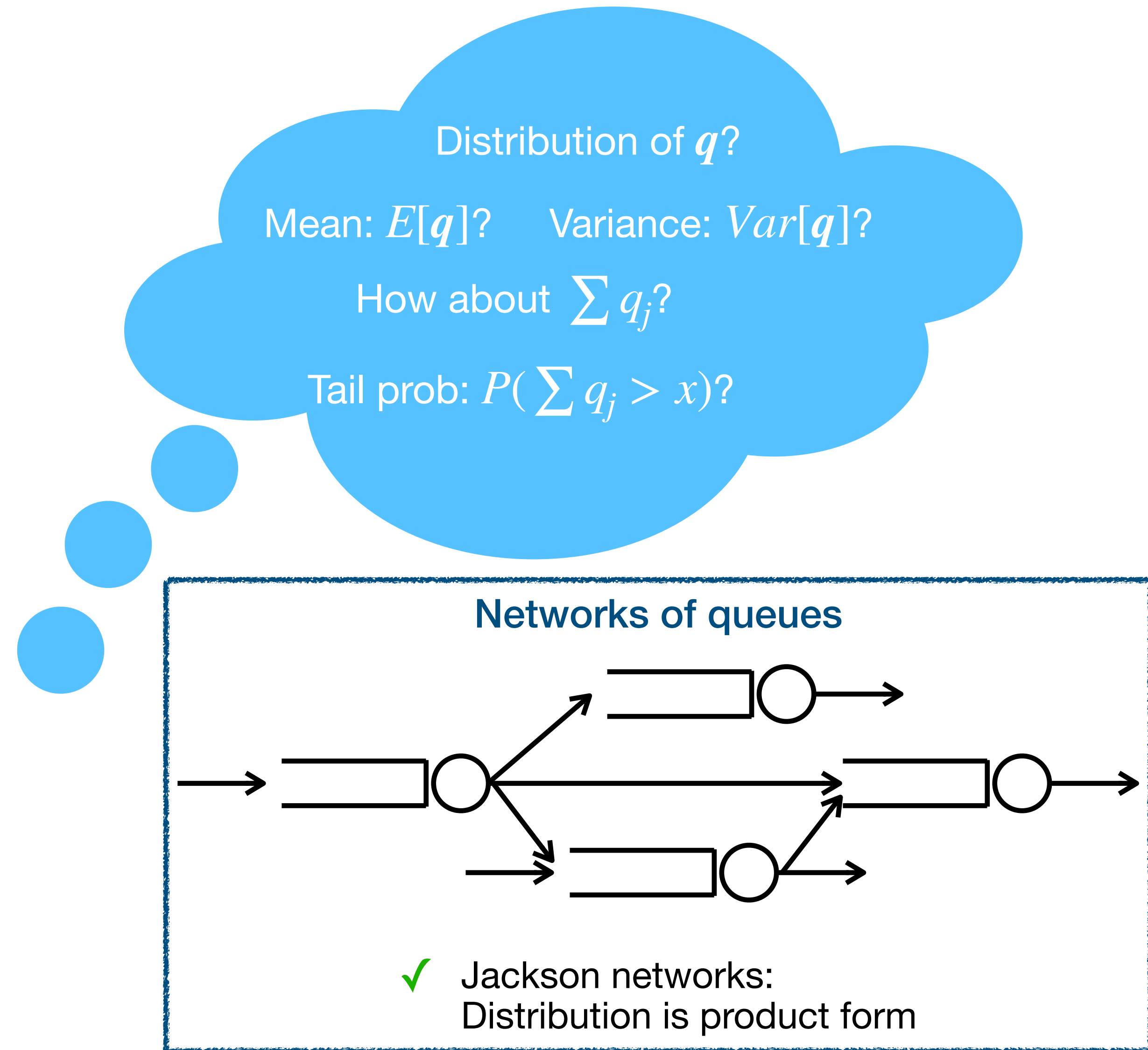
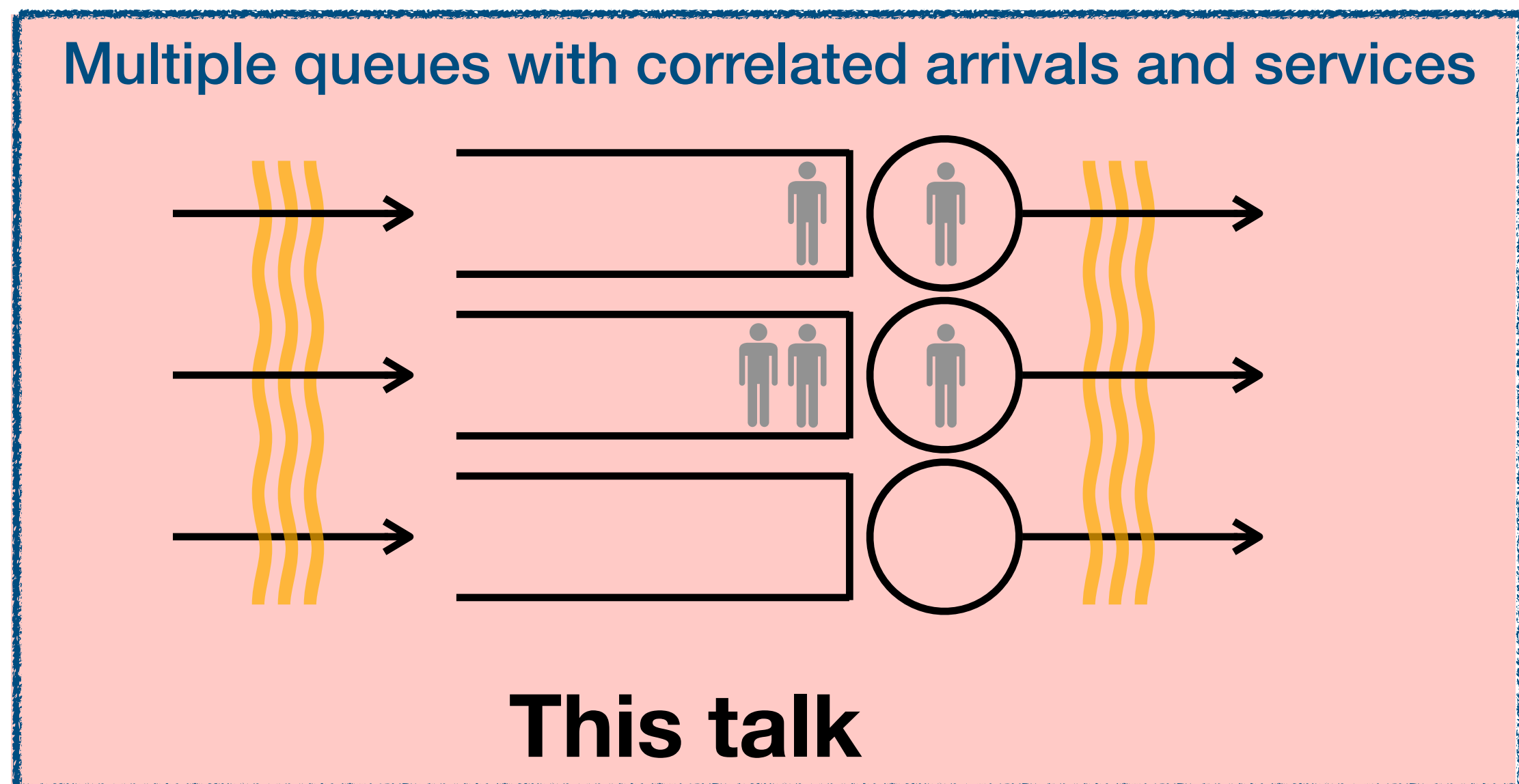
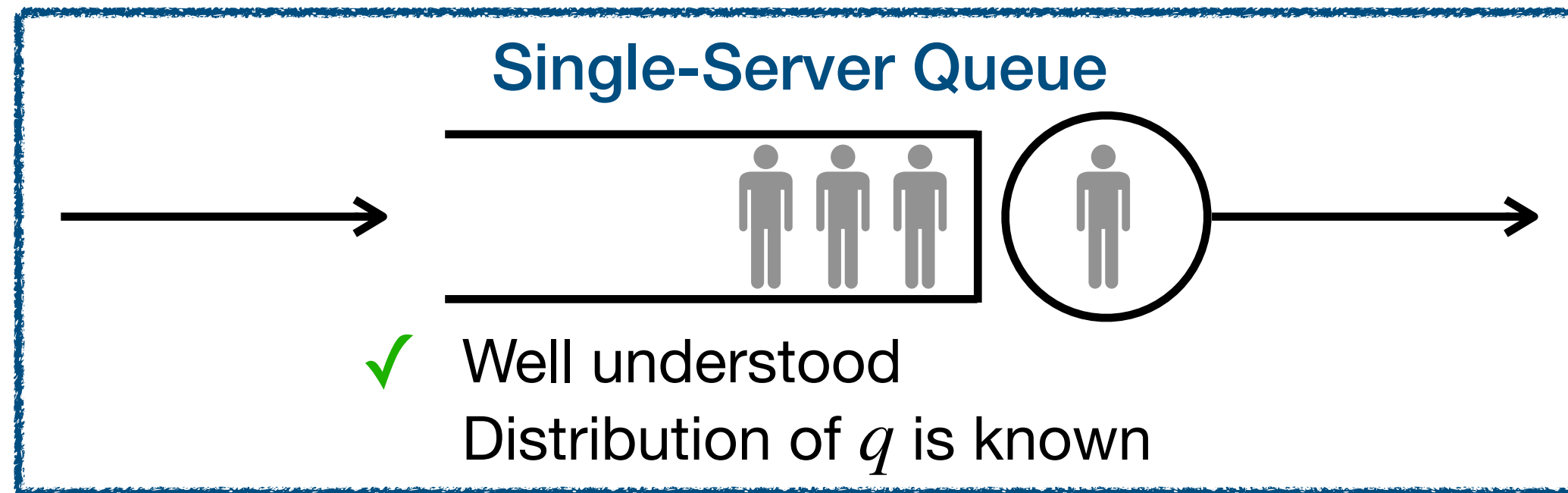
Book Now



Source: TaskRabbit website

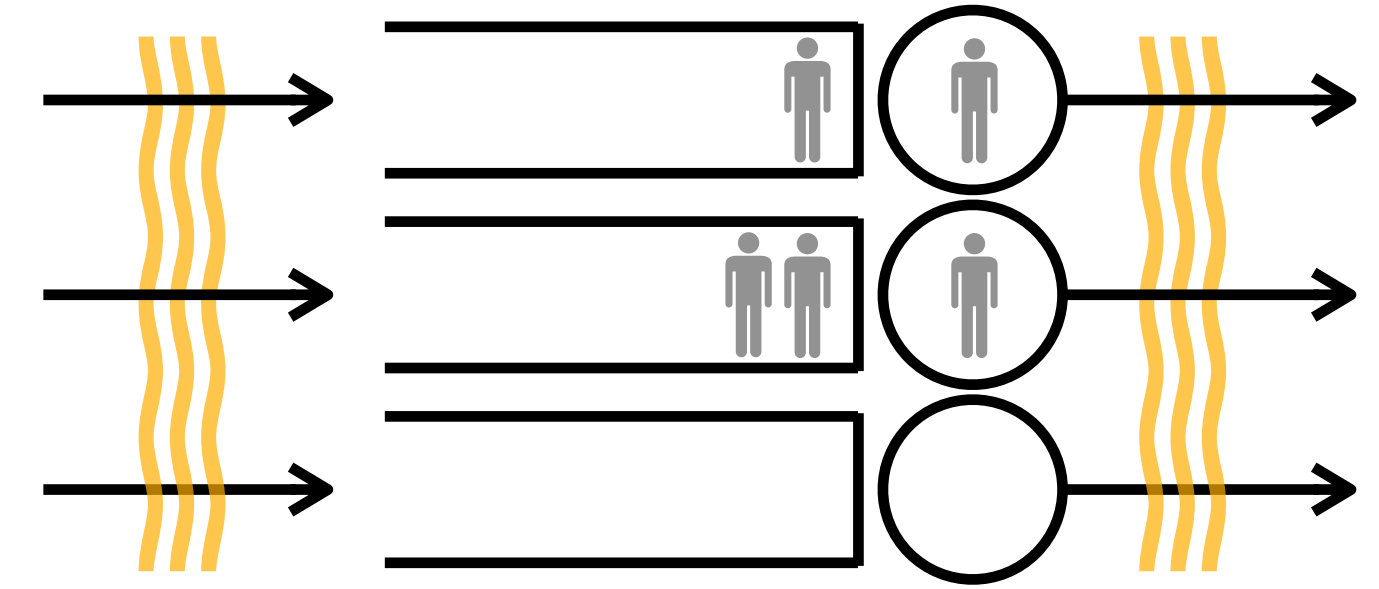


Understanding Delay



In the literature...

Asymptotic Regimes



Large Deviations:
(a.k.a. rare events)
 $\lim_{x \rightarrow \infty} P(\text{wait} > x)$

Ganesh (2004). Big Queues.
Glynn and Whitt (1994)
Puhalskii (1995)
Duffield and O'Connell (1995)
Dupuis and Ellis (1995)
...

Mean Field:
(a.k.a. many servers)
servers $\rightarrow \infty$

Vvedenskaya et.al. (1996)
Mitzenmacher (1996, 2001)
Stolyar (2017)
Ying (2017)
Mukkerjee et.al. (2018)
...

Many-Server Heavy-Traffic:
servers $\rightarrow \infty$
& Arrival rate \approx Service rate

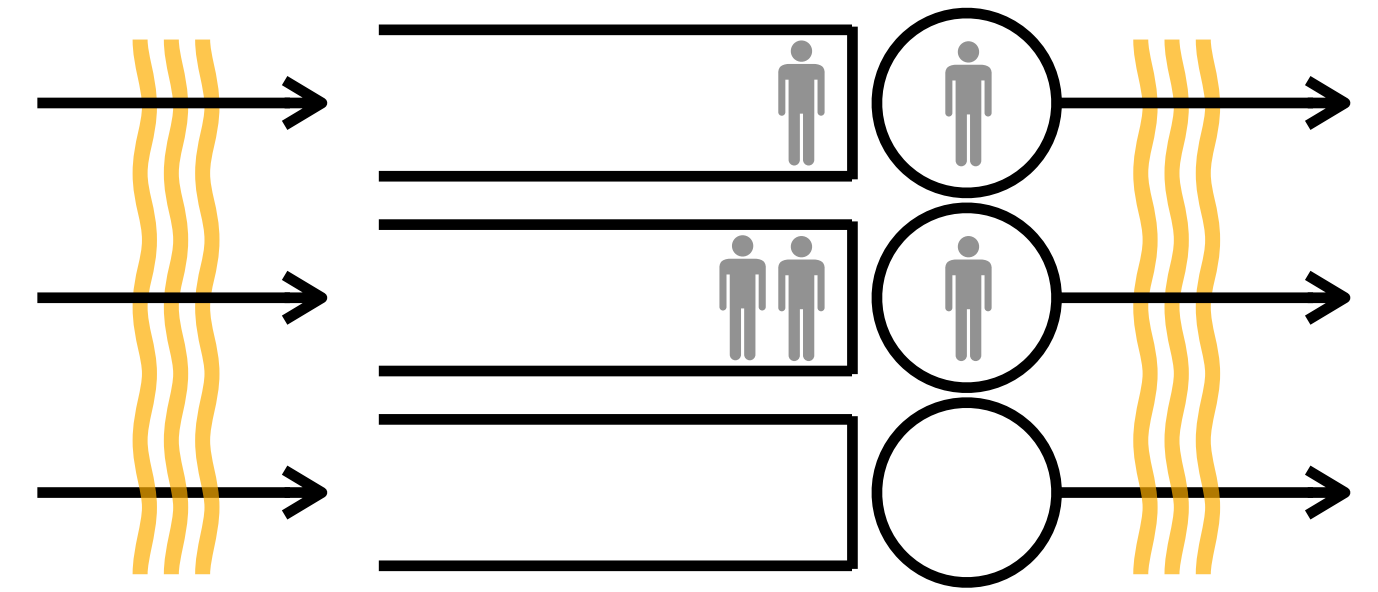
Banerjee et.al. (2019, 2020)
Liu and Ying (2019)
Braverman (2020)
Varma et.al. (2022)
HL, Maguluri (2022)
Jhunhunwala, **HL**, Maguluri (2023)
...

Heavy Traffic:
Load system to max capacity
 \Leftrightarrow Arrival rate \approx Service rate

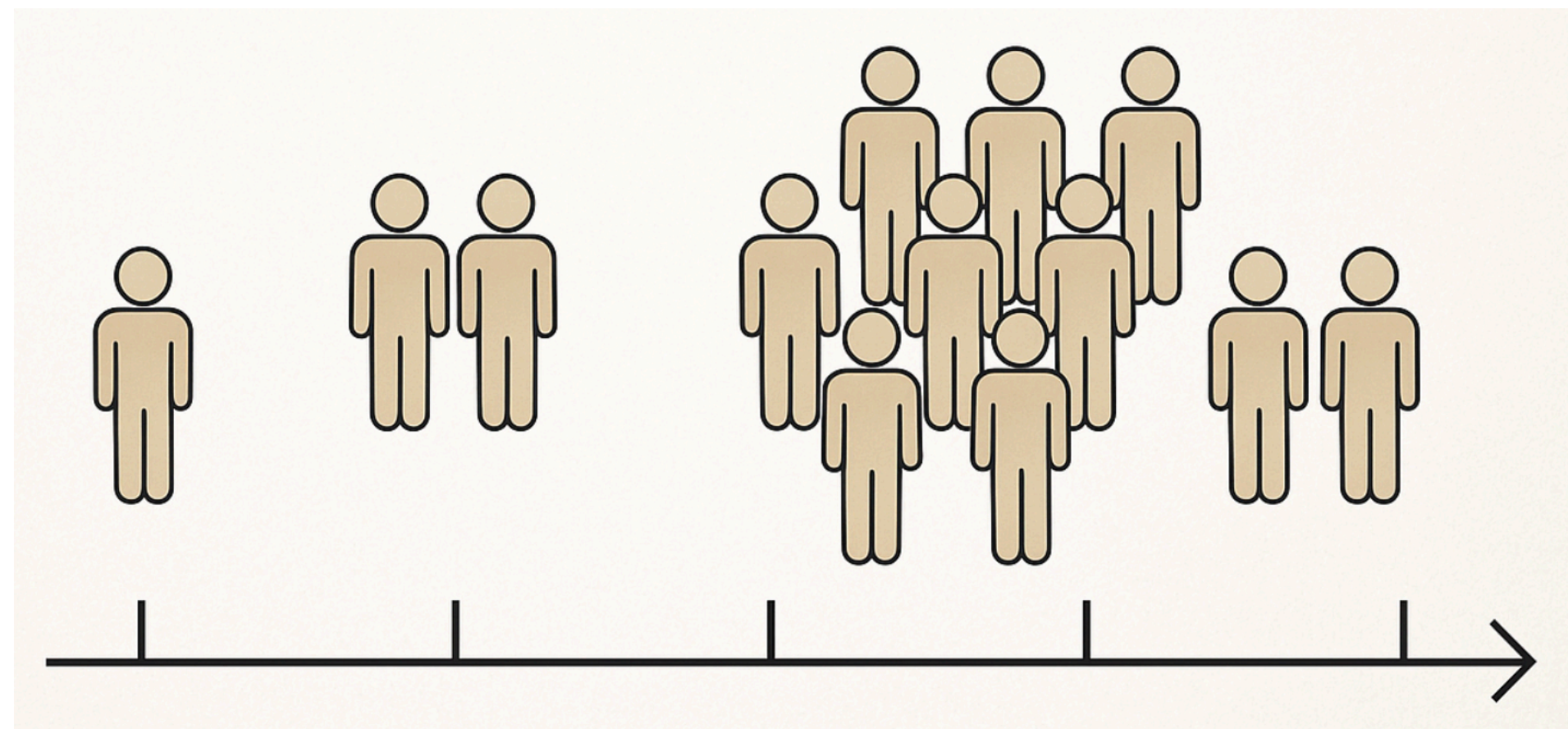
Kingman (1962); Harrison (1998)
Williams (1998, 2000)
Harrison and López (1999)
Stolyar (2004)
Gamarnik and Zeevi (2006)
Eryilmaz and Srikant (2012)
HL, Maguluri (2020, 2021)
HL et.al. (2022)
...

But all these assume constant arrival and service rates

Beyond fixed parameters



Arrival rate:



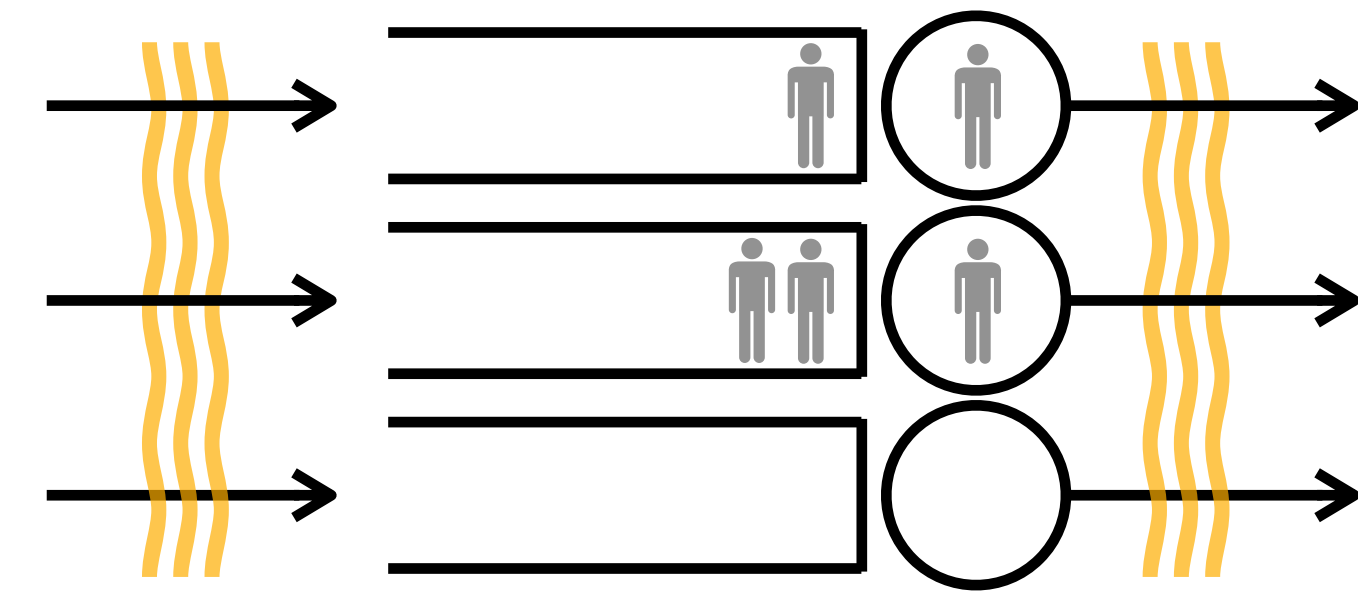
Service rate:



Working vacations



Literature (continued)



Asymptotic Regimes

Large Deviations:
(a.k.a. rare events)
 $\lim_{x \rightarrow \infty} P(\text{wait} > x)$

Glynn and Whitt (1994); Puhalskii (1995); Duffield and O'Connell (1995); Dupuis and Ellis (1995); Ganesh (2004)...

Mean Field:
(a.k.a. many servers)
servers $\rightarrow \infty$

Our contribution:

- ✓ Multiple queues and servers
- ✓ Markov-modulated arrival and service rates, countable state space
- ✓ Heavy-traffic analysis

Many-Server Heavy-Traffic
servers $\rightarrow \infty$
& Arrival rate \approx Service rate

(2022); HL, Maguluri (2022); Jhunjhunwala, HL, Maguluri (2023)...

Heavy Traffic:
Load system to max capacity
 \iff Arrival rate \approx Service rate

Kingman (1962); Harrison (1998); Williams (1998, 2000); Harrison and López (1999); Stolyar (2004); Gamarnik and Zeevi (2006); Eryilmaz and Srikant (2012); HL, Maguluri (2020, 2021); HL et.al. (2022)...

Constant arrival and service rates

Markov Modulated Queues

- Burman, Smith (1986): Markov/M/1 queue in light and heavy traffic. Approximation of mean.
- Prabhu, Zhu (1989): Busy period and mean workload in a Markov/Markov/1 queue
- ... (1991): Markov/G/1 queue analysis
- ... (2005): Single-server queue with Markov-modulated service times
- ...

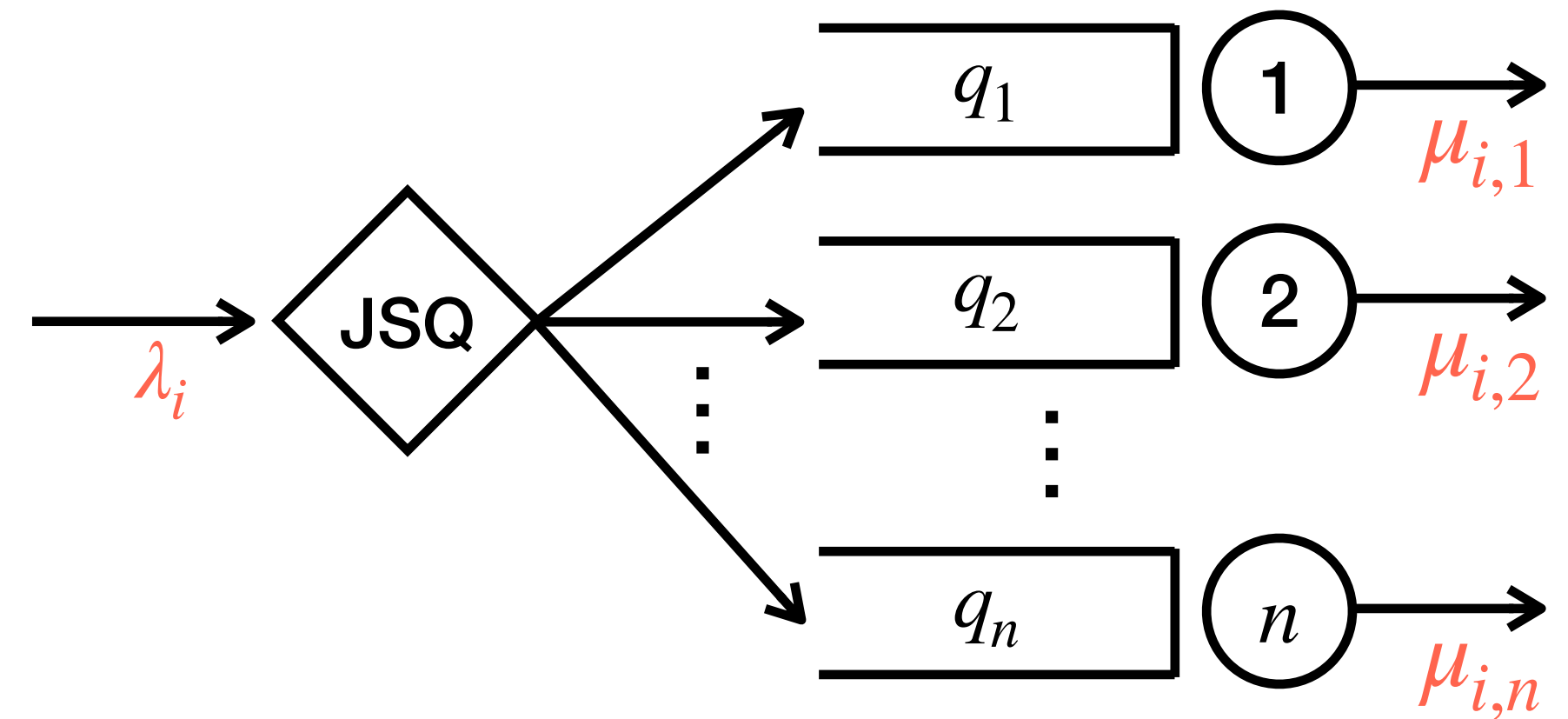
↑ Single queue with finite-state Markov modulation

- Maguluri, Mou (2024): Input-queued switch with Markov-modulated arrivals

Multiple queues, but constant service

JSQ Model

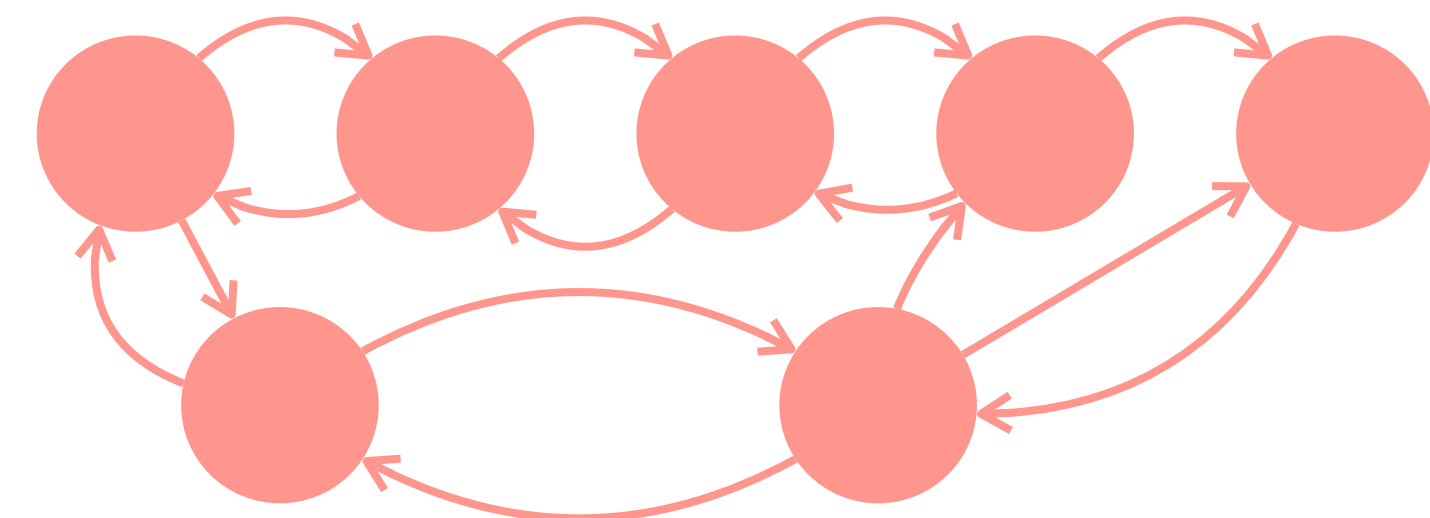
- Load balancing under Join-the-Shortest-Queue (JSQ)
- Exponential inter-arrival and service times
- Heterogeneous servers
- Arrival and service rate are **Markov-Modulated**



Assumptions on $\{Z(t)\}_t$

- Countable state space
- Stationary distribution exists
- $\lambda_{\max} < \infty$ and $\mu_{\max} < \infty$

$Z(t) = i \sim$ Markov chain



$$\rho_\pi = \frac{\mathbb{E}[\lambda_i | i \sim \pi]}{\sum_j \mathbb{E}[\mu_{ij} | i \sim \pi]}, \quad \epsilon = 1 - \rho_\pi$$

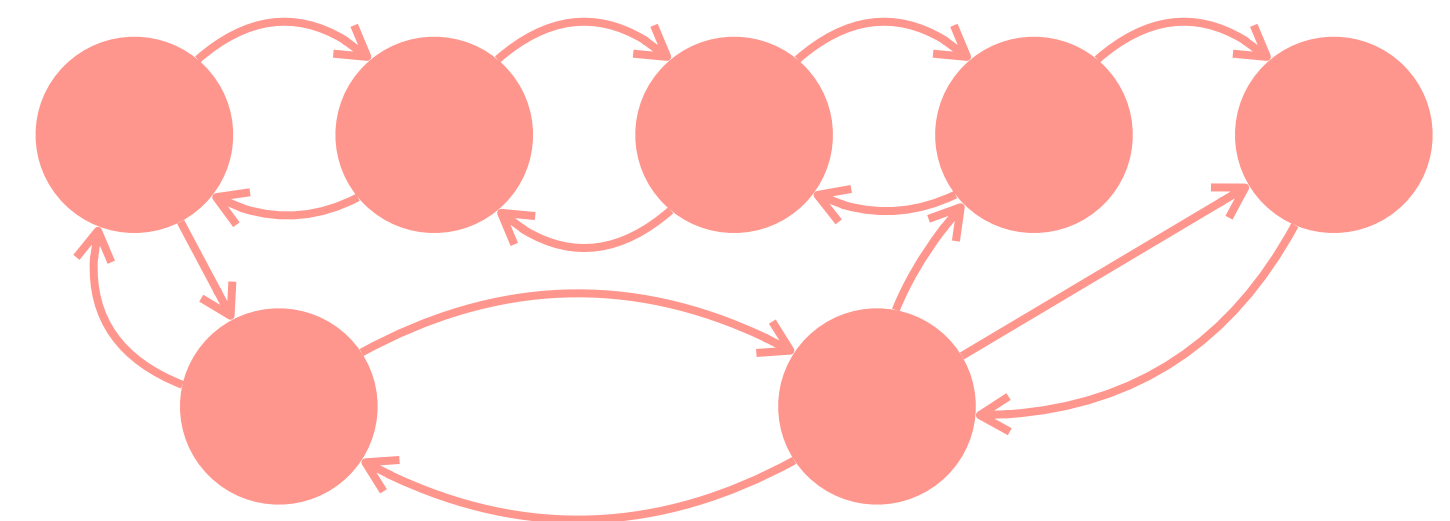
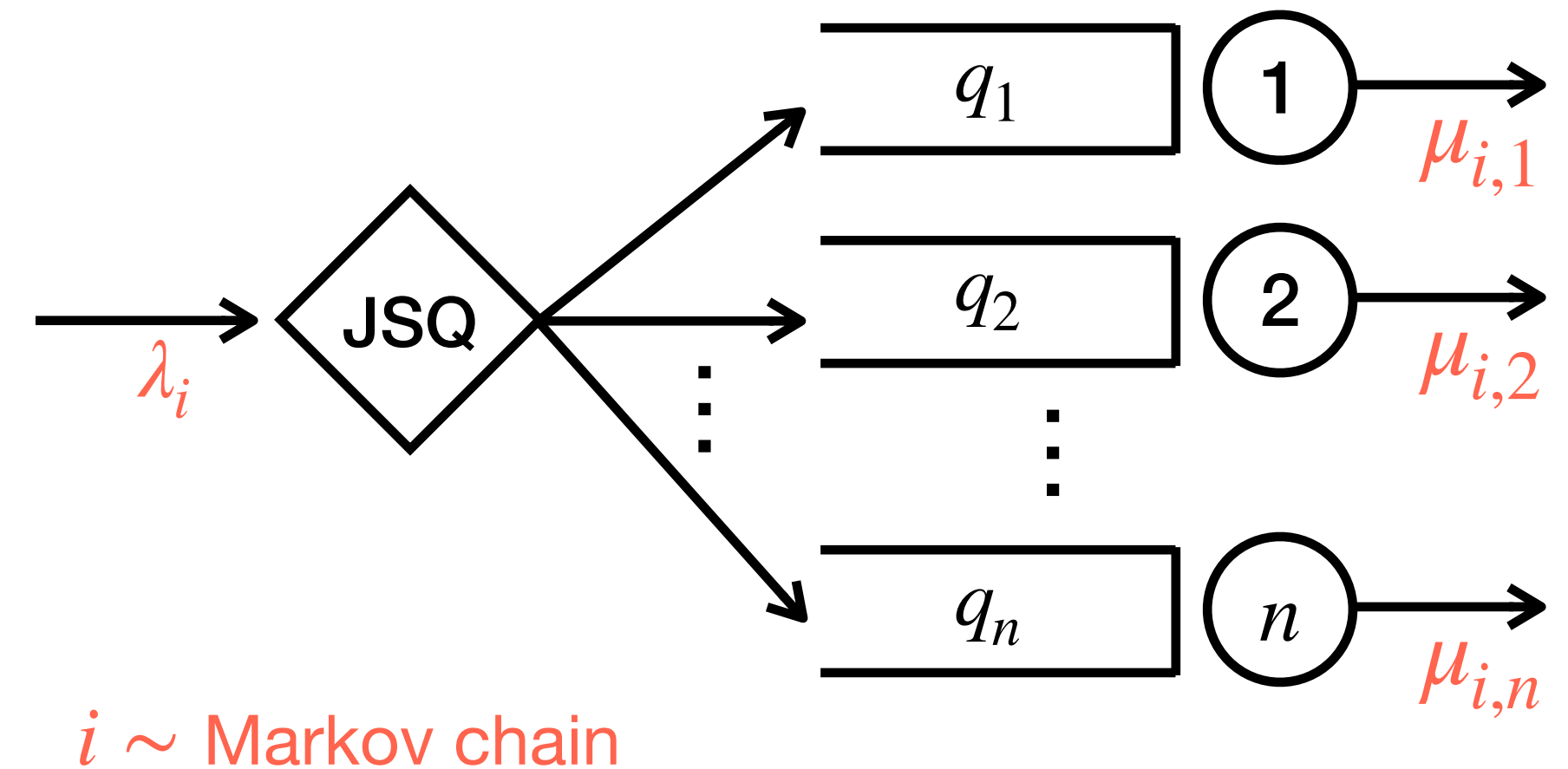
Asymptotic Distribution of Queue Lengths

Theorem [HL, Grosf '25]:

If $\lambda_i > 0$ is large enough for each i , then as $\epsilon \downarrow 0$

$$\epsilon q \implies \text{Exp} \left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma} \right)$$

\approx Variance of arrival and service times in steady state



$$\rho_\pi = \frac{\mathbb{E}[\lambda_i | i \sim \pi]}{\sum_j \mathbb{E}[\mu_{ij} | i \sim \pi]}, \quad \epsilon = 1 - \rho_\pi$$

Proof Sketch

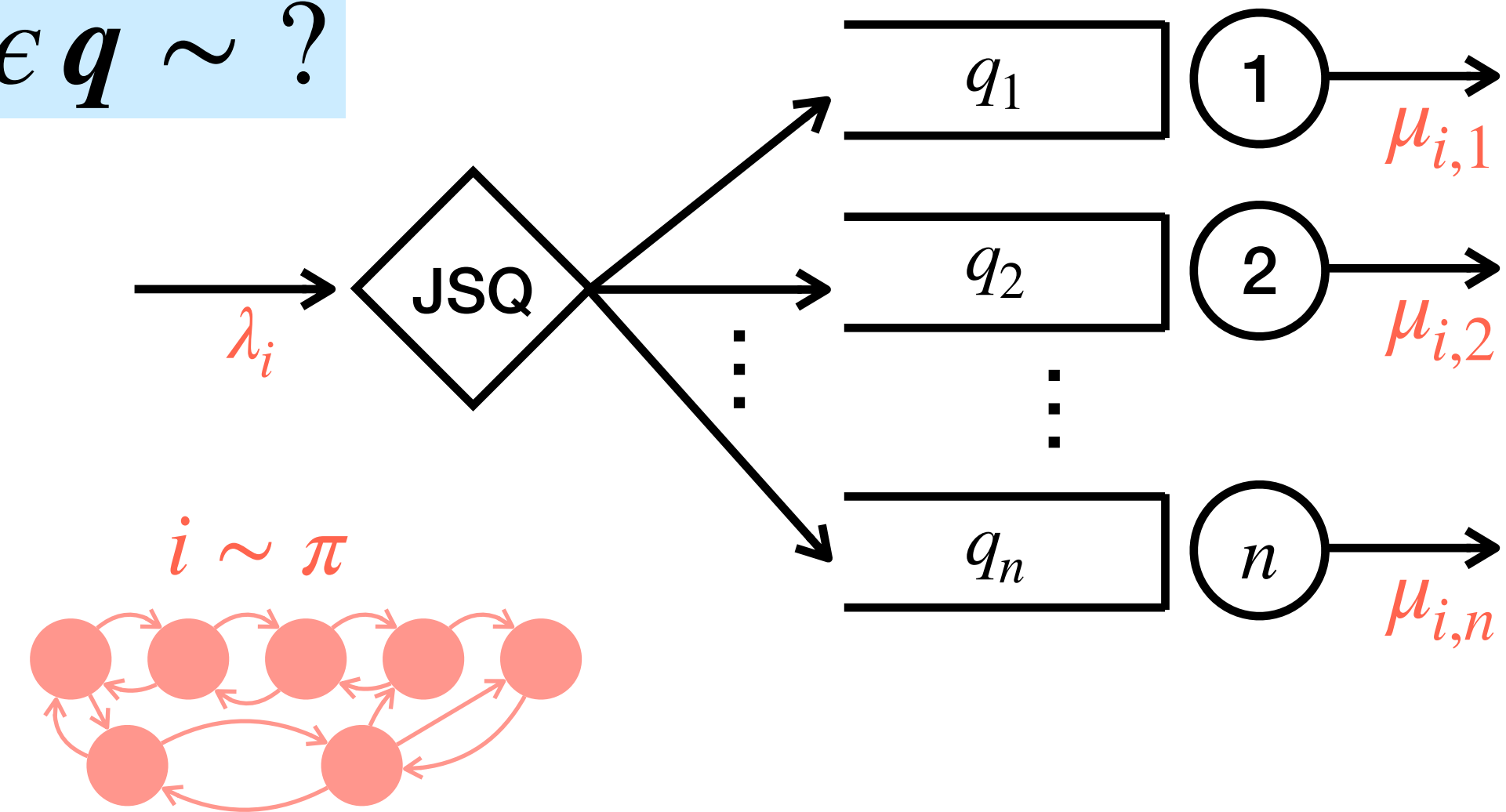
Theorem [HL, Grosfot '25]:

If $\lambda_i > 0$ is large enough for each i , then as $\epsilon \downarrow 0$

$$\epsilon q \implies \text{Exp} \left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma} \right)$$

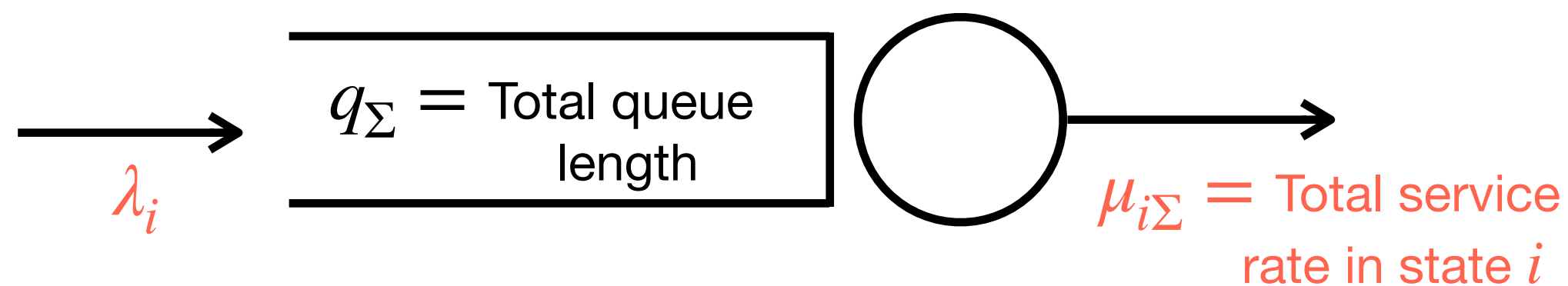
\approx Variance of arrival and service times in steady state

$$\epsilon q \sim ?$$



Step 1: State Space Collapse (SSC)

$q \approx \left(\frac{q_\Sigma}{n} \right) \mathbf{1}$, so study the following single-server queue:



Step 2: Asymptotic distribution

$$\epsilon q_\Sigma \sim ?$$

We use:

- Transform Method
- Poisson Equation

State Space Collapse (SSC)

Proposition [HL, Grosf '25]:

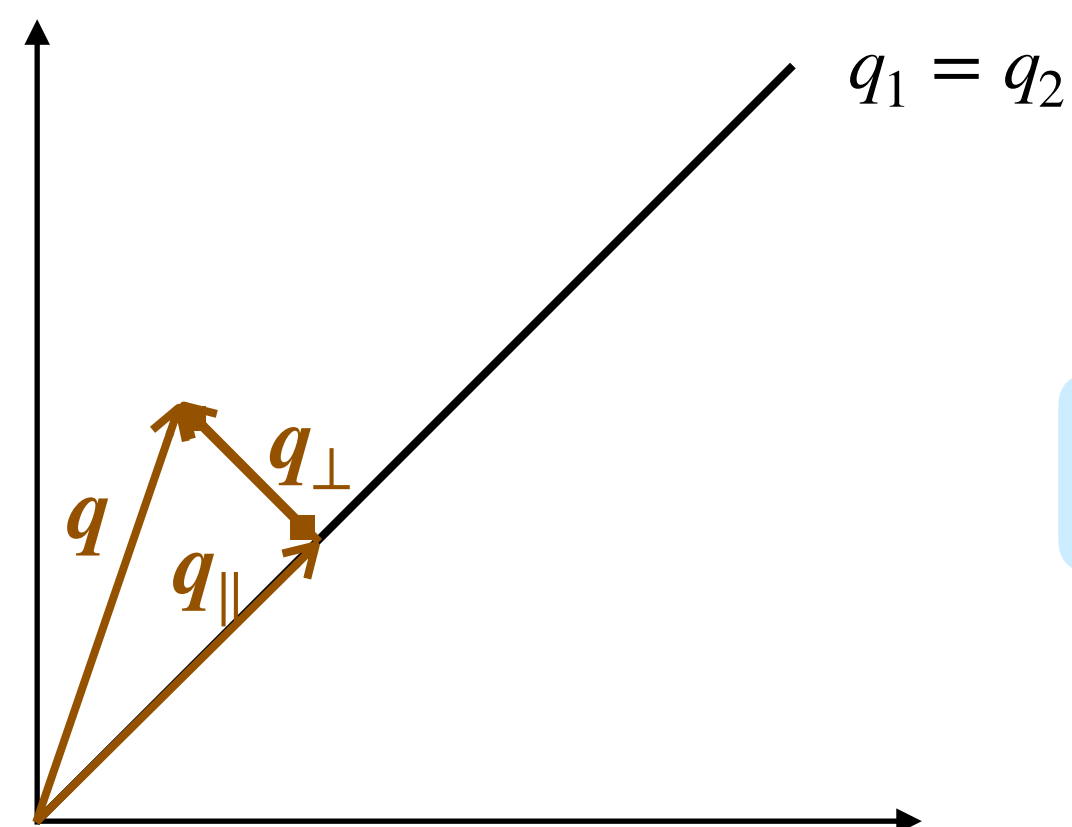
q_{\parallel} := projection of q on $\mathbf{1}$

q_{\perp} := $q - q_{\parallel}$

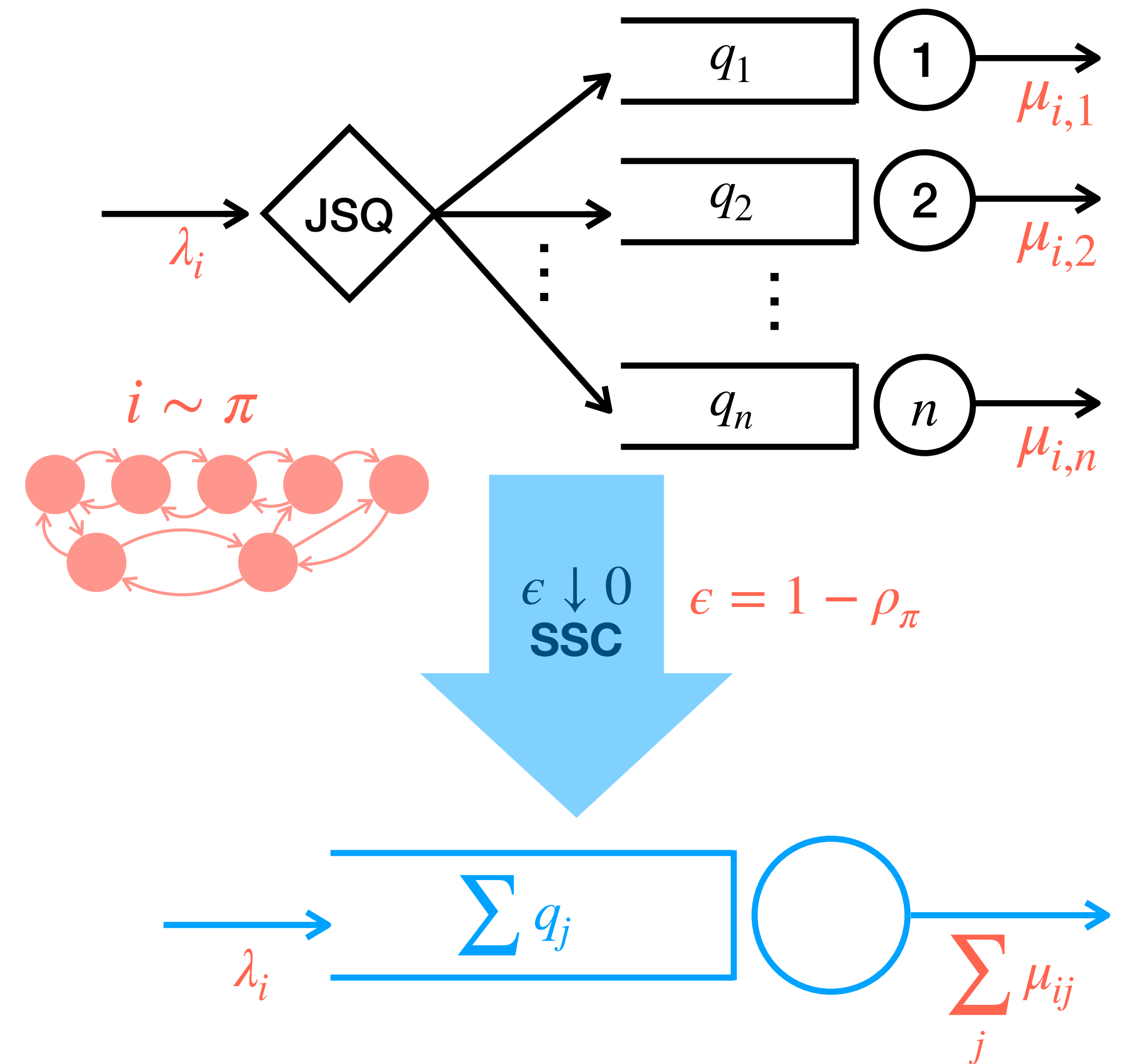
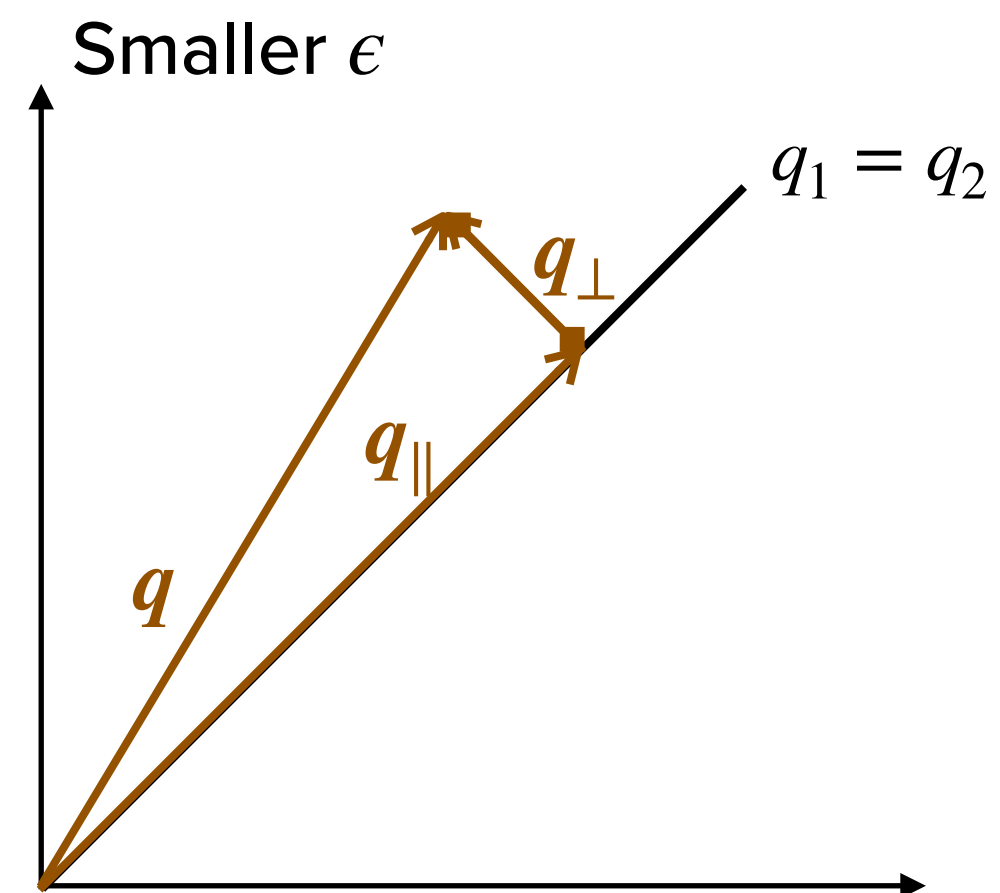
If $\lambda_i > 0$ is large enough for each i , for each $m \in \mathbb{N}$,

$$\mathbb{E} [\|q_{\perp}\|^m] \leq M_m$$

Proof: Drift analysis



$$q \approx q_{\parallel}$$



State Space Collapse (SSC)

Proposition [HL, Grosf '25]:

$q_{\parallel} :=$ projection of q on $\mathbf{1}$

$q_{\perp} := q - q_{\parallel}$

If $\lambda_i > 0$ is large enough for each i , for each $m \in \mathbb{N}$,

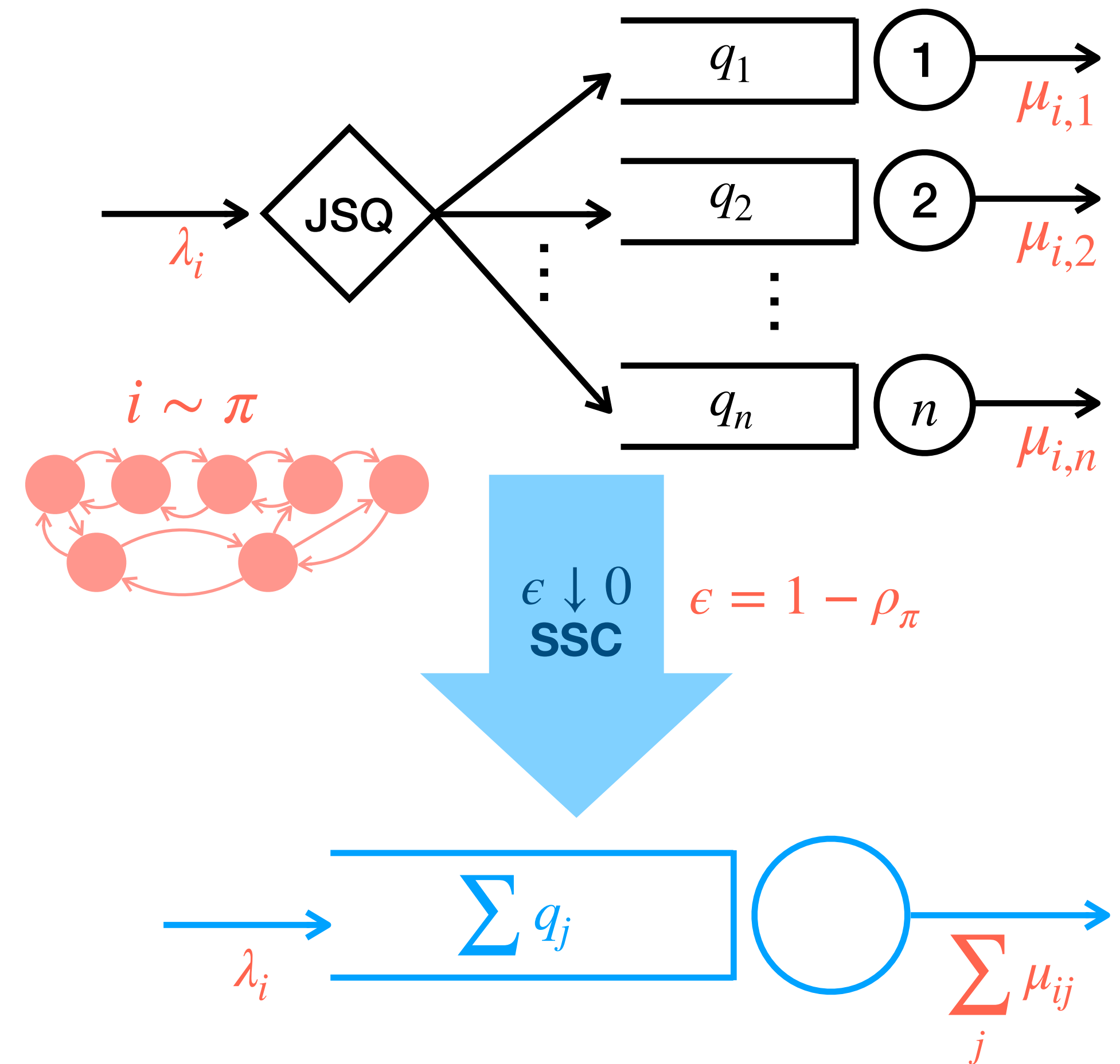
$$\mathbb{E} [\|q_{\perp}\|^m] \leq M_m$$

What is λ_i large enough?

$$\lambda_i > \mu_{i\Sigma} - n \min_j \mu_{ij}$$

When does JSQ balance the queue lengths?

- ✓ For homogeneous servers, $\lambda_i > 0$ is enough
- ✓ “Enough” depends on heterogeneity of servers
- ✓ Arrivals need to cover range of service rates
- ✓ JSQ only balances q with *enough arrivals*



Proof Sketch

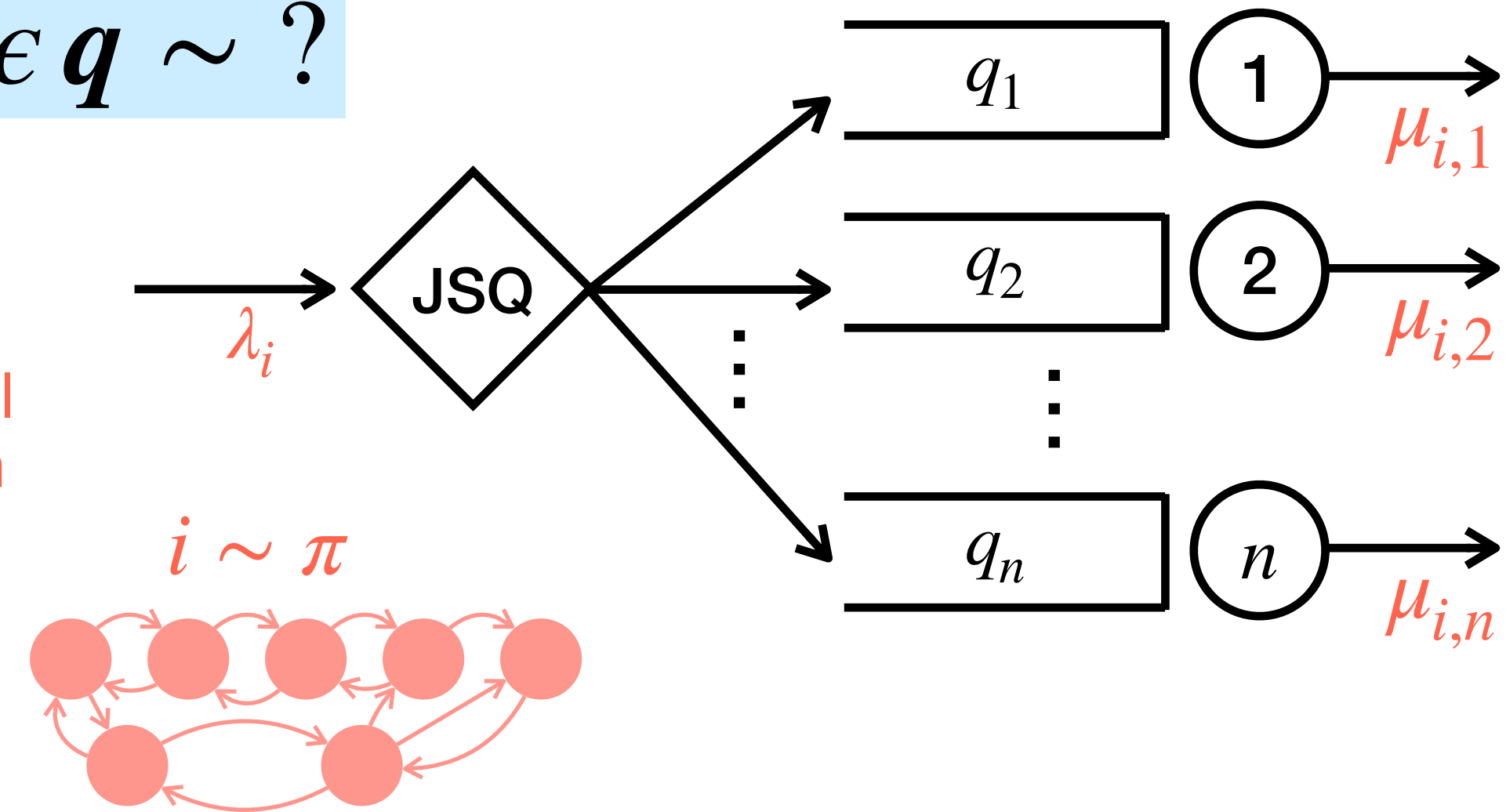
Theorem [HL, Grosfot '25]:

If $\lambda_i > 0$ is large enough for each i , then as $\epsilon \downarrow 0$

$$\epsilon q \implies \text{Exp} \left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma} \right)$$

\approx Variance of arrival and service times in steady state

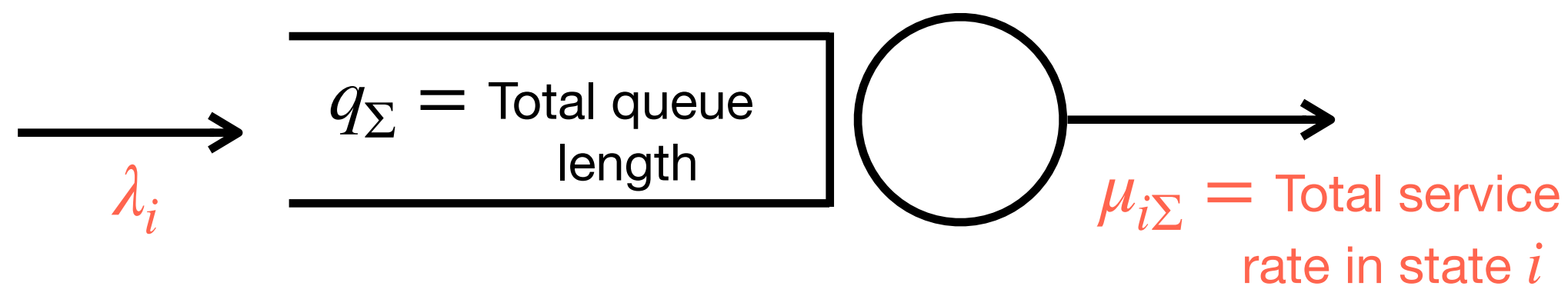
$$\epsilon q \sim ?$$



Step 1: State Space Collapse (SSC)



$q \approx \left(\frac{q_\Sigma}{n} \right) \mathbf{1}$, so study the following single-server queue:



Step 2: Asymptotic distribution

$$\epsilon q_\Sigma \sim ?$$

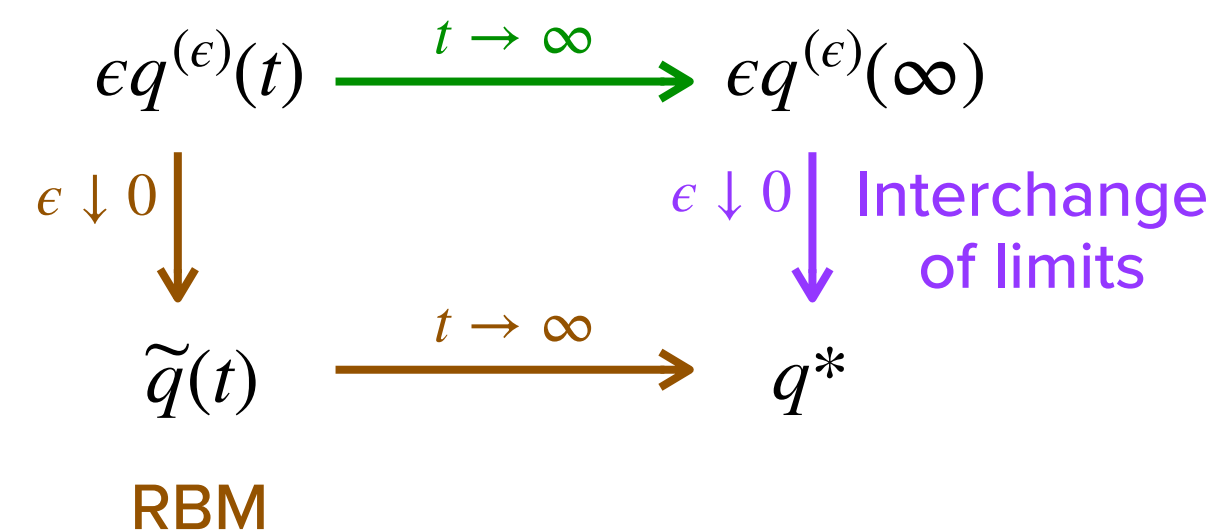
We use:

- Transform Method
- Poisson Equation

Heavy-Traffic Analysis in the Literature

Diffusion Limits Approach

- Most popular
- Introduced by Kingman (1962)
- Show convergence in distribution of queue length scaled heavy-traffic parameter $1 - \rho$



Direct Methods

- **Stein's method**
Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*
- **Drift Method**
Inductively compute $\mathbb{E} [(\epsilon q)^k]$ for $k \in \mathbb{N}$
- **BAR approach**
Analyze $e^{\epsilon q(\infty)}$ and bound jumps appropriately
- **Transform Methods**
Analyze drift of $e^{\epsilon q^{(\epsilon)}(\infty)}$ and directly compute $\mathbb{E} [e^{\epsilon q^{(\epsilon)}(\infty)}]$
 - ✓ Tractable analysis
 - ✓ Compute distribution
 - ✓ Obtain tail bounds

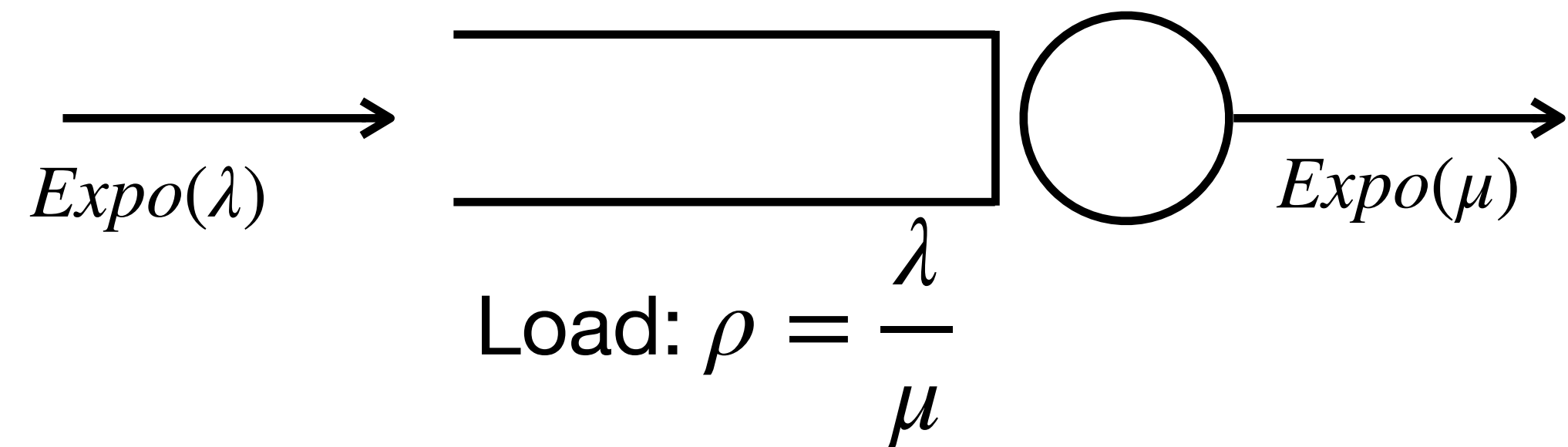
Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

$$\varphi_\theta(q) = e^{\theta q}, \theta \in \mathbb{R}$$



$$\begin{aligned} \Delta\varphi_\theta(q) &= \overset{\text{Arrival}}{\lambda(e^{\theta(q+1)} - e^{\theta q})} + \overset{\text{Departure}}{\mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})} \\ &= e^{\theta q} (e^{-\theta} - 1) (\mu - \lambda e^\theta) - \mu (e^{-\theta} - 1) e^{\theta q} 1_{\{q=0\}} \\ &= e^{\theta q} (e^{-\theta} - 1) (\mu - \lambda e^\theta) - \mu (e^{-\theta} - 1) 1_{\{q=0\}} \end{aligned}$$

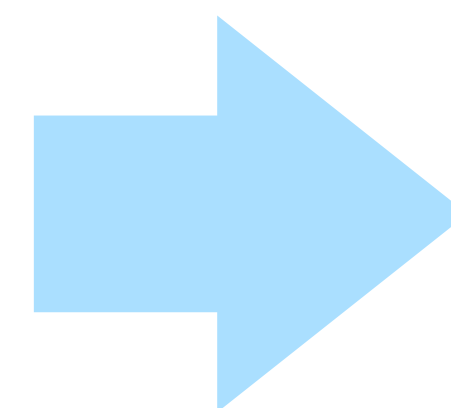


Drift:

$$\Delta\varphi_\theta(q) = \mathbb{E} \left[e^{\theta q(t+)} - e^{\theta q} \mid q(t) = q \right]$$

Step 2: Set drift to zero

$$\mathbb{E} \left[e^{\theta q} \right] = \frac{\mathbb{P}[q = 0]}{1 - \rho e^\theta}$$

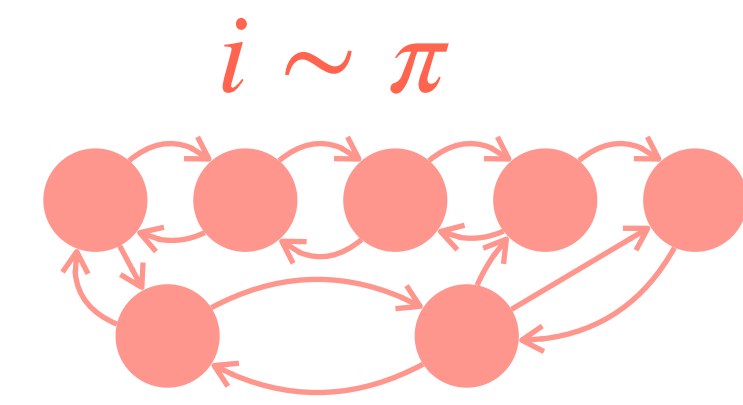


Step 3: Set $\theta = 0$

$$\mathbb{P}[q = 0] = 1 - \rho$$

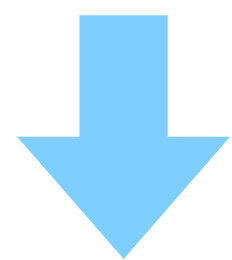
$$\begin{aligned} \mathbb{E} \left[e^{\theta q} \right] &= \frac{1 - \rho}{1 - \rho e^\theta} \\ \iff q &\sim \text{Geometric}(1 - \rho) \end{aligned}$$

Transform Method Markov-Mod JSQ



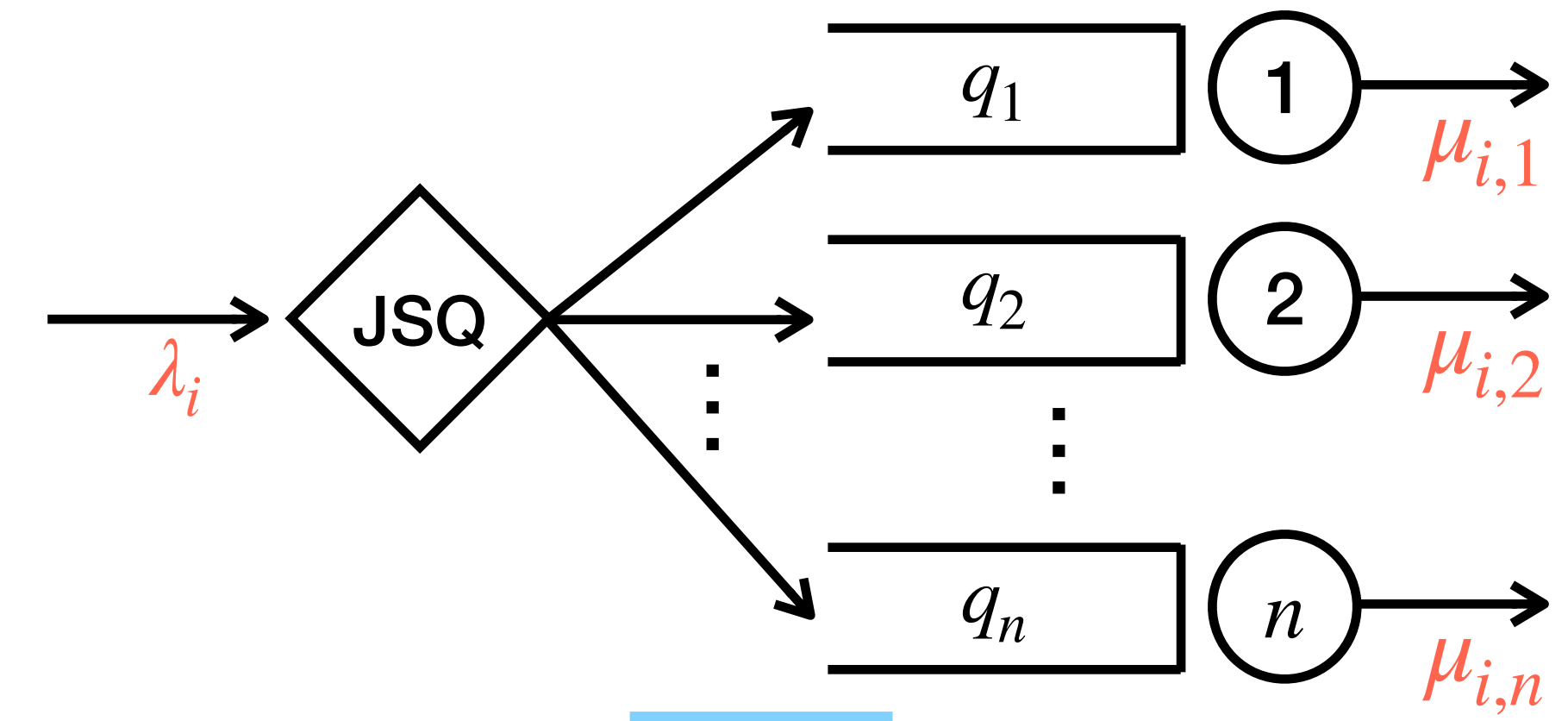
Step 1: Drift of exponential test function

$$\varphi_s(i, \mathbf{q}) = e^{-s\epsilon q_\Sigma}, s > 0$$



$$\Delta\varphi_s(i, \mathbf{q}) = (e^{-s\epsilon} - 1) e^{-s\epsilon q_\Sigma} (\lambda_i - \mu_{i\Sigma} e^{s\epsilon})$$

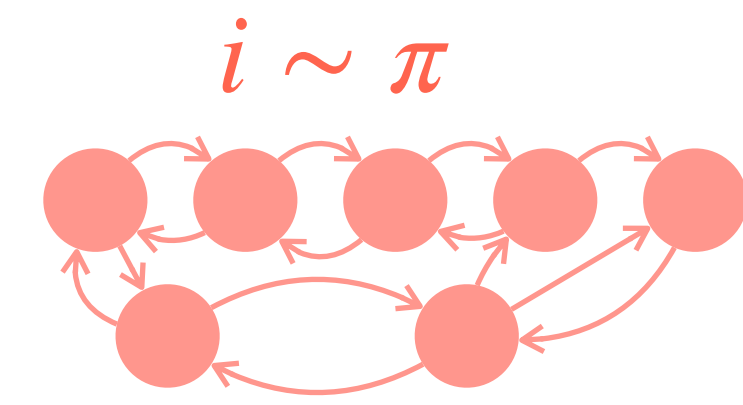
$$\begin{aligned}
 & - (e^{-s\epsilon} - 1) e^{-s\epsilon q_\Sigma} \left(\sum_j \mu_{ij} 1_{\{q_j=0\}} \right) \\
 & \approx \sum_j \mu_{ij} e^{-senq_j} 1_{\{q_j=0\}} \quad (\text{SSC}) \\
 & = \sum_j \mu_{ij} 1_{\{q_j=0\}}
 \end{aligned}$$



$$\epsilon := 1 - \rho_\pi$$

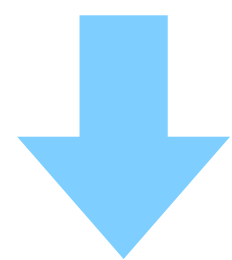


Transform Method Markov-Mod JSQ



Step 1: Drift of exponential test function

$$\varphi_s(i, \mathbf{q}) = e^{-s\epsilon q_\Sigma}, s > 0$$



$$\Delta\varphi_s(i, \mathbf{q}) \approx (e^{-s\epsilon} - 1) e^{-s\epsilon q_\Sigma} (\lambda_i - \mu_{i\Sigma} e^{s\epsilon})$$

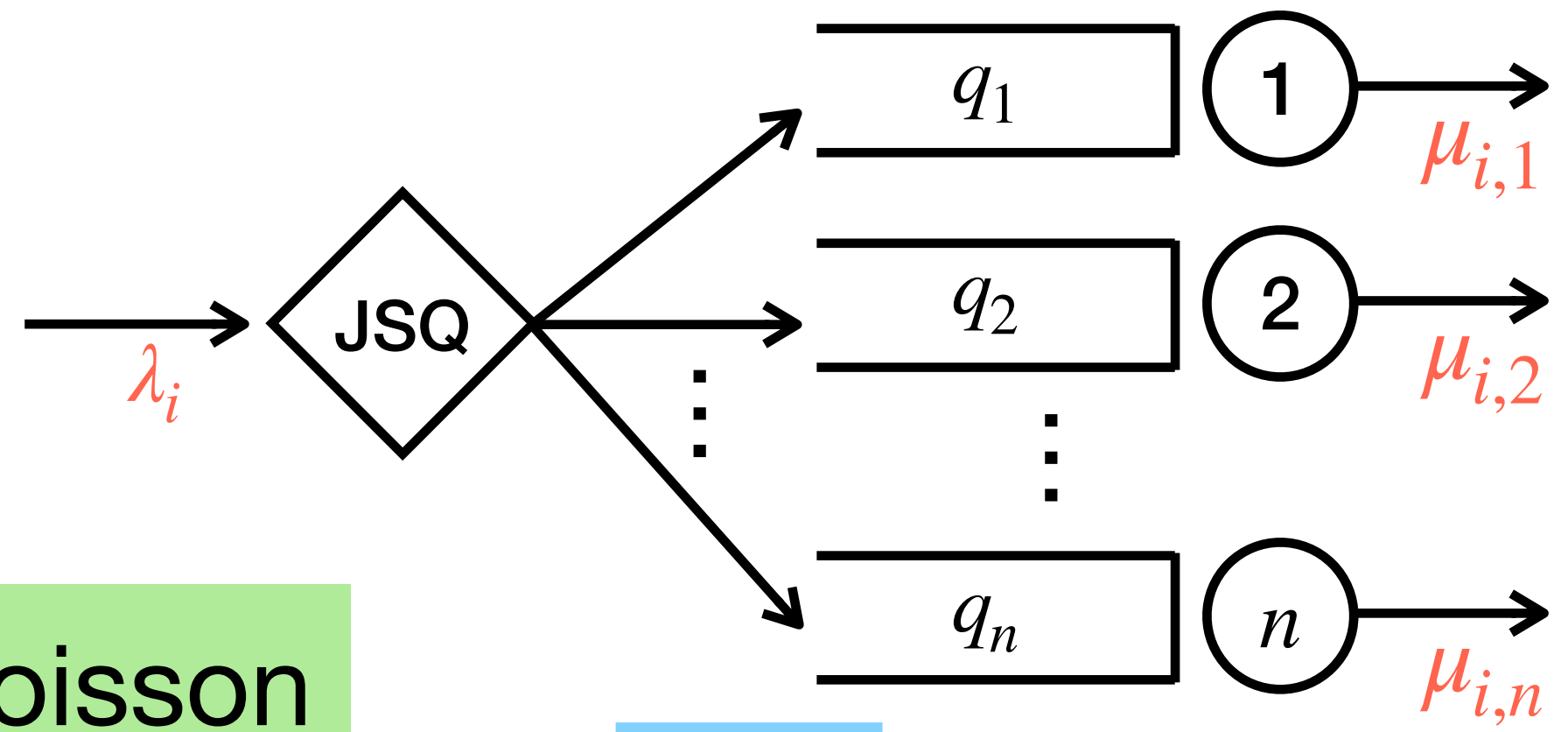
Not independent!

How to compute $\mathbb{E}[\Delta\varphi_s(i, \mathbf{q})]$?

We use the Poisson equation!

$$-(1 - e^{s\epsilon}) \left(\sum_j \mu_{ij} 1_{\{q_j=0\}} \right)$$

$$\mathbb{E} \left[\sum_j \mu_{ij} 1_{\{q_j=0\}} \right] = \mu_{i\Sigma} \epsilon$$



$$\epsilon := 1 - \rho_\pi$$

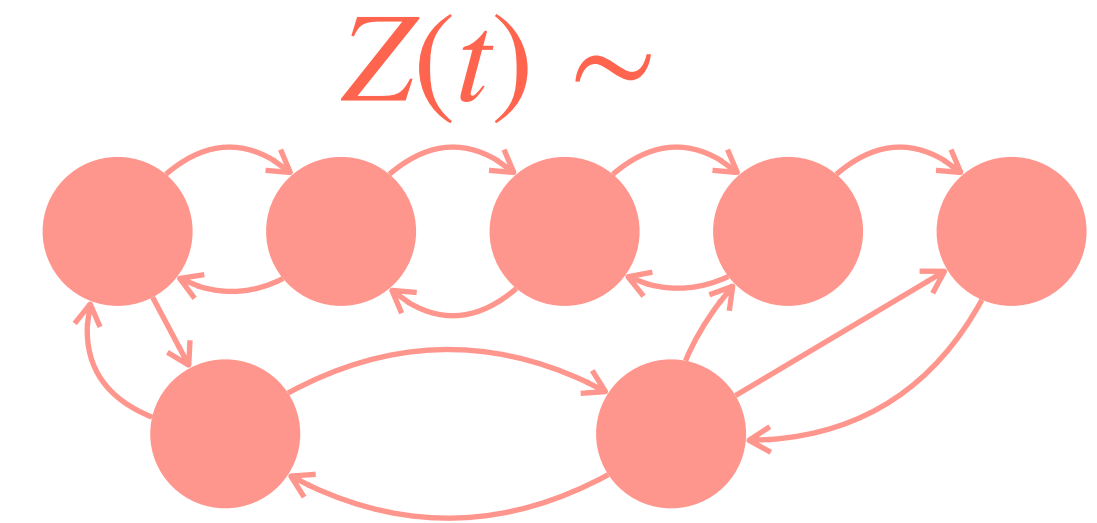


Poisson Equation

Poisson equation:

Let $\{Z(t)\}_t$ be a CTMC with countable state space \mathcal{L} and transition rates $\alpha_{i,i'}$. Consider a function $f: \mathcal{L} \rightarrow \mathbb{R}$ and let $\bar{f} = \mathbb{E}[f(Z)]$. Then, there exists a function $V_f: \mathcal{L} \rightarrow \mathbb{R}$ such that

$$V_f(i) = \frac{f(i) - \bar{f}}{\alpha_i} + \sum_{i' \neq i} \frac{\alpha_{ii'}}{\alpha_i} V_f(i')$$



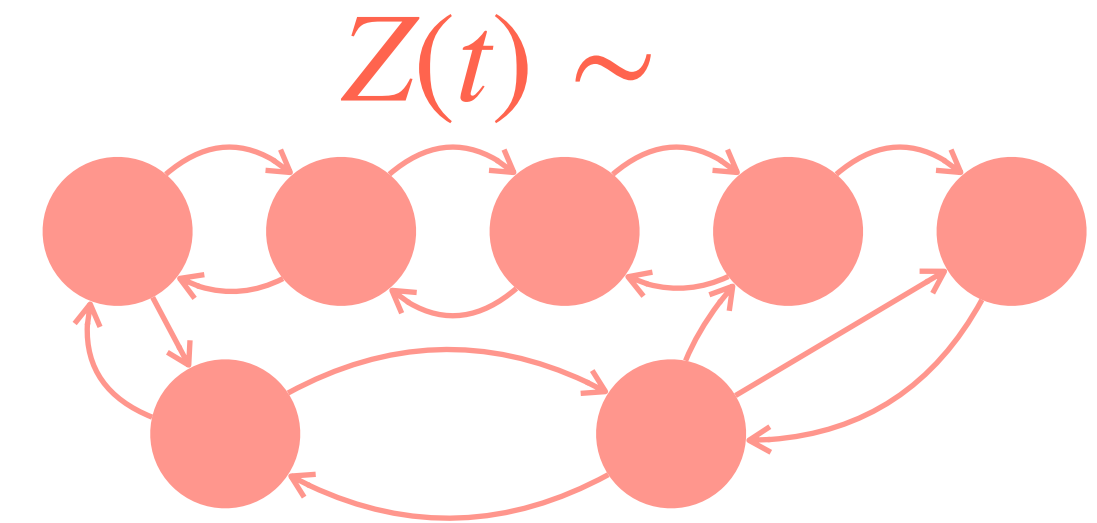
$$\Leftrightarrow \underbrace{\sum_{i' \neq i} \alpha_{ii'} [V_f(i') - V_f(i)]}_{\text{Drift of solution } V_f(i)} = \underbrace{\bar{f} - f(i)}_{\text{Difference between average and } f(i)}$$

Poisson Equation for Transform Method

Poisson equation:

Let $\{Z(t)\}_t$ be a CTMC with countable state space \mathcal{L} and transition rates $\alpha_{i,i'}$. Consider a function $f: \mathcal{L} \rightarrow \mathbb{R}$ and let $\bar{f} = \mathbb{E}[f(Z)]$. Then, there exists a function $V_f: \mathcal{L} \rightarrow \mathbb{R}$ such that

$$V_f(i) = \frac{f(i) - \bar{f}}{\alpha_i} + \sum_{i' \neq i} \frac{\alpha_{ii'}}{\alpha_i} V_f(i')$$



Theorem [HL, Grosf '25]:

For any function $f: \mathcal{L} \rightarrow \mathbb{R}$ such that $V_f(i)$ exists and $\mathbb{E} \left[|V_f(i)|^{1+\eta} \right] \leq C$ for some $\eta > 0$,

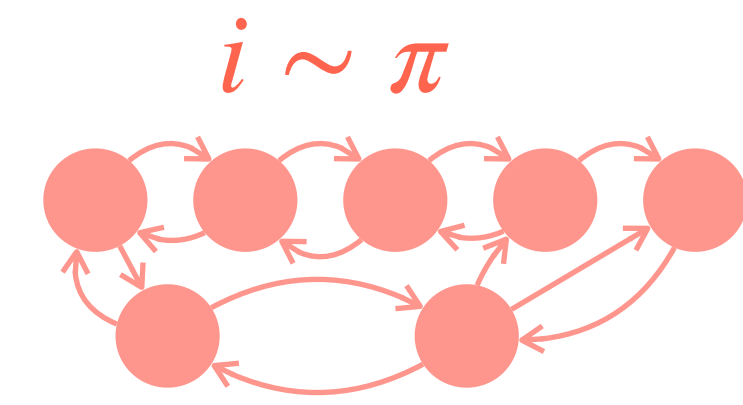
$$\text{Cov} \left(e^{-s\epsilon q_\Sigma}, f(i) \right) = \mathbb{E} \left[e^{-s\epsilon q_\Sigma} f(i) \right] - \mathbb{E} \left[e^{-s\epsilon q_\Sigma} \right] \bar{f} = (e^{-s\epsilon} - 1) \mathbb{E} \left[e^{-s\epsilon q_\Sigma} V_f(i) (\lambda_i - \mu_{i\Sigma}) \right] + O \left(\epsilon^{2 - \frac{1}{1+\eta}} \right)$$

✓ Product of expectations

Solution to
Poisson equation

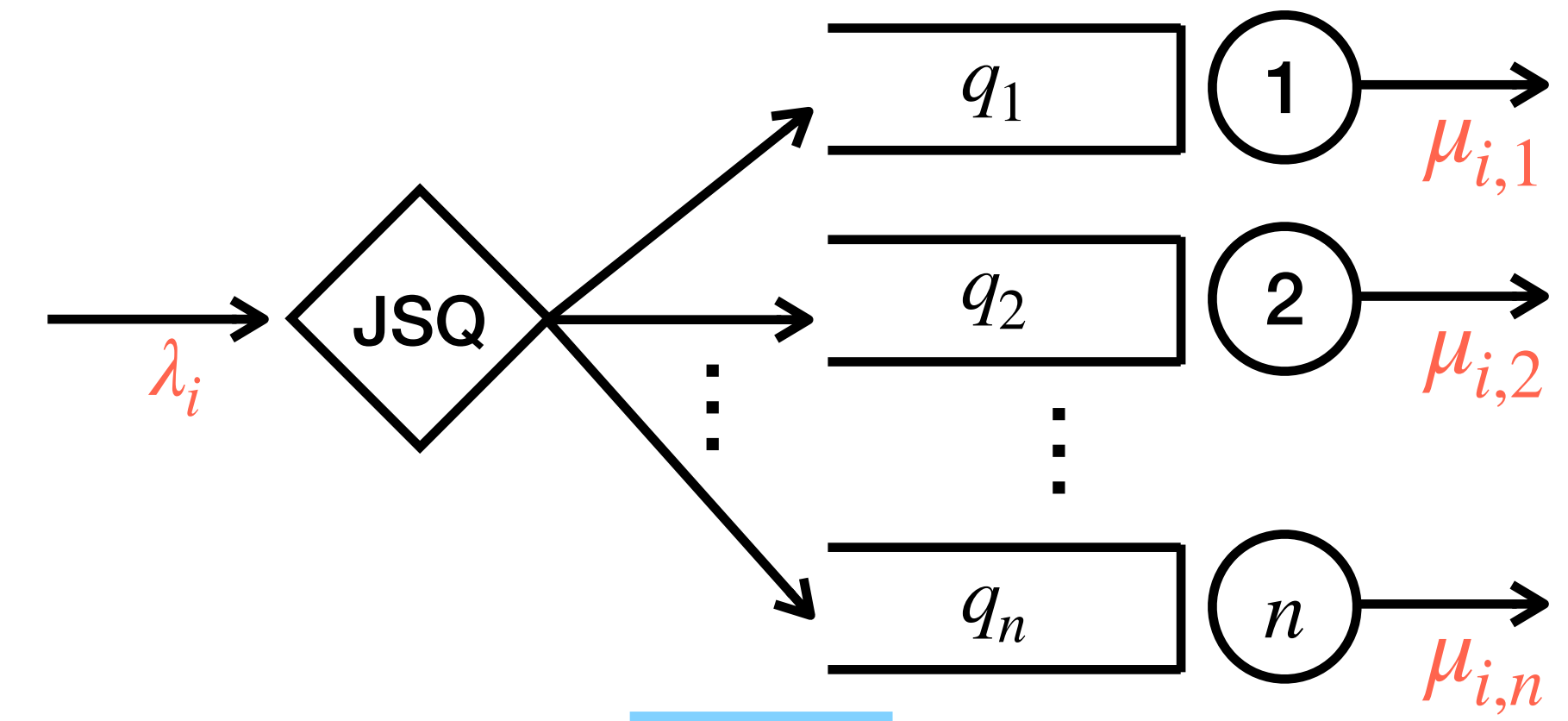
Error term

Transform Method Markov-Mod JSQ



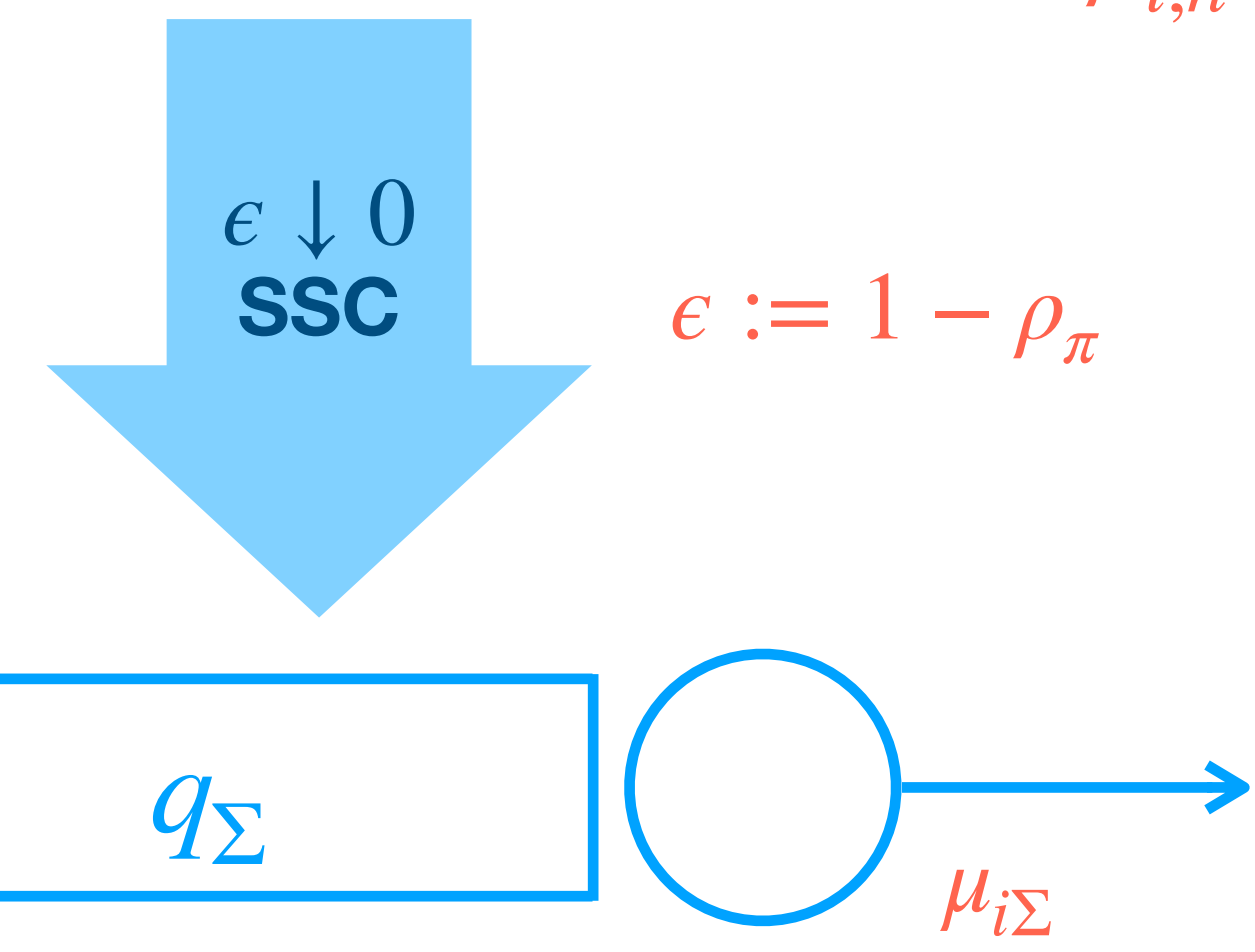
Step 1: Drift of exponential test function

$$\Delta\varphi_s(i, \mathbf{q}) \approx (e^{-s\epsilon} - 1) e^{-s\epsilon q_\Sigma} (\lambda_i - \mu_{i\Sigma} e^{s\epsilon}) - (1 - e^{s\epsilon}) \left(\sum_j \mu_{ij} 1_{\{q_j=0\}} \right)$$



Step 2: Expectation under steady state

$$\underbrace{\mathbb{E} \left[e^{-s\epsilon q_\Sigma} (e^{s\epsilon} \mu_{i\Sigma} - \lambda_i) \right]}_{\text{Step 3: Our theorem!}} = e^{s\epsilon} \underbrace{\mathbb{E} \left[\sum_j \mu_{ij} 1_{\{q_j=0\}} \right]}_{= \mu_\Sigma \epsilon} + O(\epsilon^2)$$



Applying the Poisson Equation

Set drift to zero: $\mathbb{E}[e^{-s\epsilon q_\Sigma} (e^{s\epsilon} \mu_{i\Sigma} - \lambda_i)] = e^{s\epsilon} \mu_\Sigma \epsilon + O(\epsilon^2)$

$$= \mathbb{E}[e^{-s\epsilon q_\Sigma} (\mu_{i\Sigma} - \lambda_i)] + s\epsilon \mathbb{E}[e^{-s\epsilon q_\Sigma} \mu_{i\Sigma}] + O(\epsilon^2)$$

Poisson equation theorem on lhs with $f(i) = h(i) := \mu_{i\Sigma} - \lambda_i$ and $f(i) = \ell(i) := \mu_{i\Sigma}$:

$$LHS = \mathbb{E}[e^{-s\epsilon q_\Sigma}] (1 + s)\epsilon \mu_\Sigma + (1 - e^{-s\epsilon}) \mathbb{E}[e^{-s\epsilon q_\Sigma} V_h(i) (\mu_{i\Sigma} - \lambda_i)] + O(\epsilon^{2 - \frac{1}{1+\eta}})$$

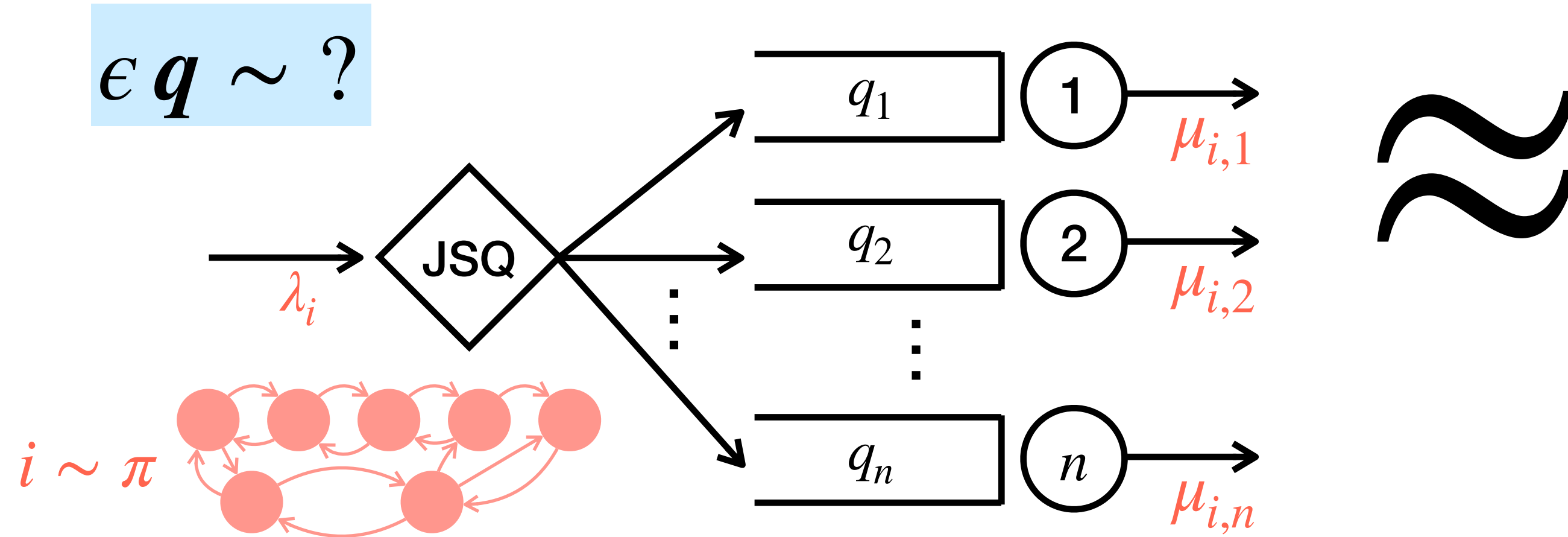
Poisson equation theorem on rhs with $f(i) = k(i) := V_h(i) (\mu_{i\Sigma} - \lambda_i)$:

$$\mathbb{E}[e^{-s\epsilon q_\Sigma} V_h(i) (\mu_{i\Sigma} - \lambda_i)] = \mathbb{E}[e^{-s\epsilon q_\Sigma}] \mathbb{E}[k(i)] + O(\epsilon)$$

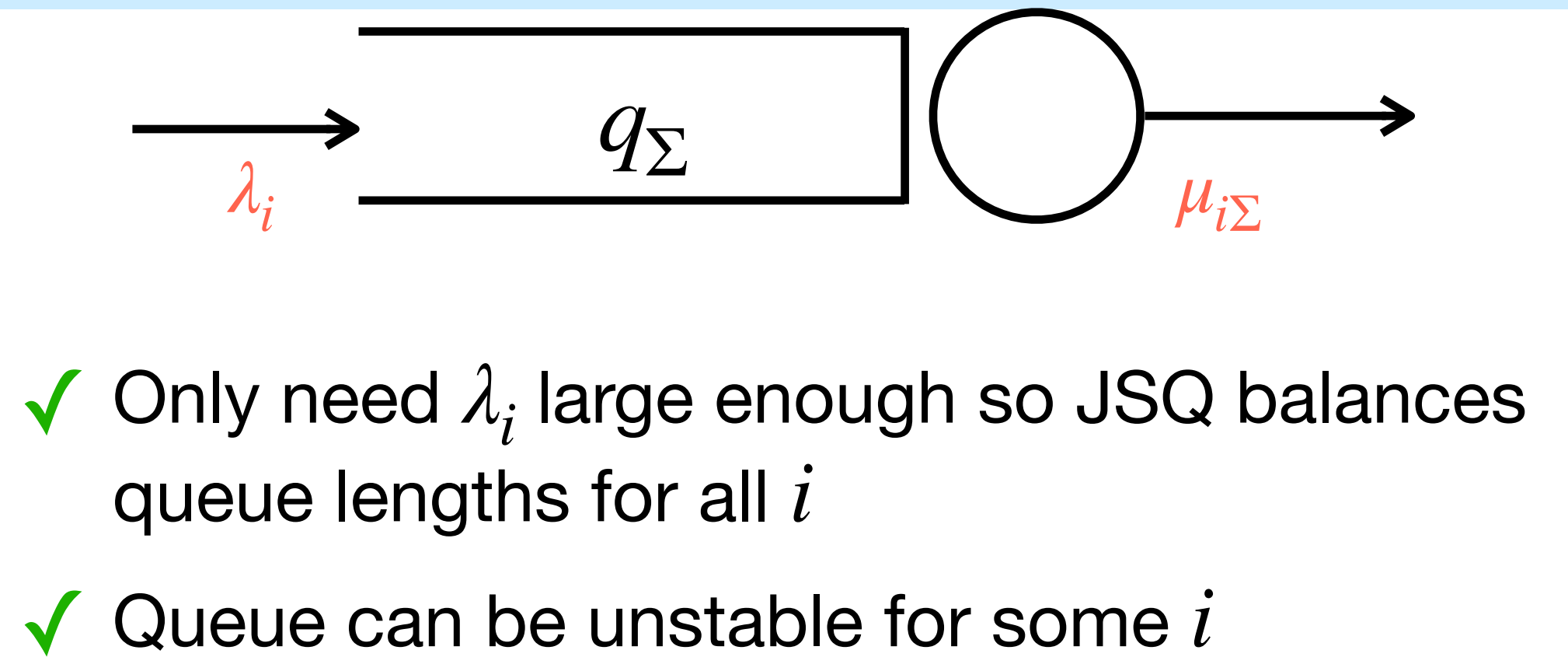
Then, put everything together. ■

Key Takeaways

Thanks! Questions?



Step 1: State Space Collapse



Step 2: Asymptotic distribution

Transform Method:
 $\varphi_s(i, \mathbf{q}) = e^{-s\epsilon q_\Sigma}, \mathbb{E}[\Delta\varphi_s(i, \mathbf{q})] = 0$
 + Poisson Equation

Theorem [HL, Grosf '25]:

$$\mathbb{E} [e^{-s\epsilon q_\Sigma}] = \frac{1}{1 + s \left(1 + \frac{\mathbb{E}[k(i)]}{\mu_\Sigma} \right)} + O \left(\epsilon^{2 - \frac{1}{1+\eta}} \right)$$

with $k(i) = V_h(i)(\mu_{i\Sigma} - \lambda_i)$ and $h(i) = \mu_{i\Sigma} - \lambda_i$