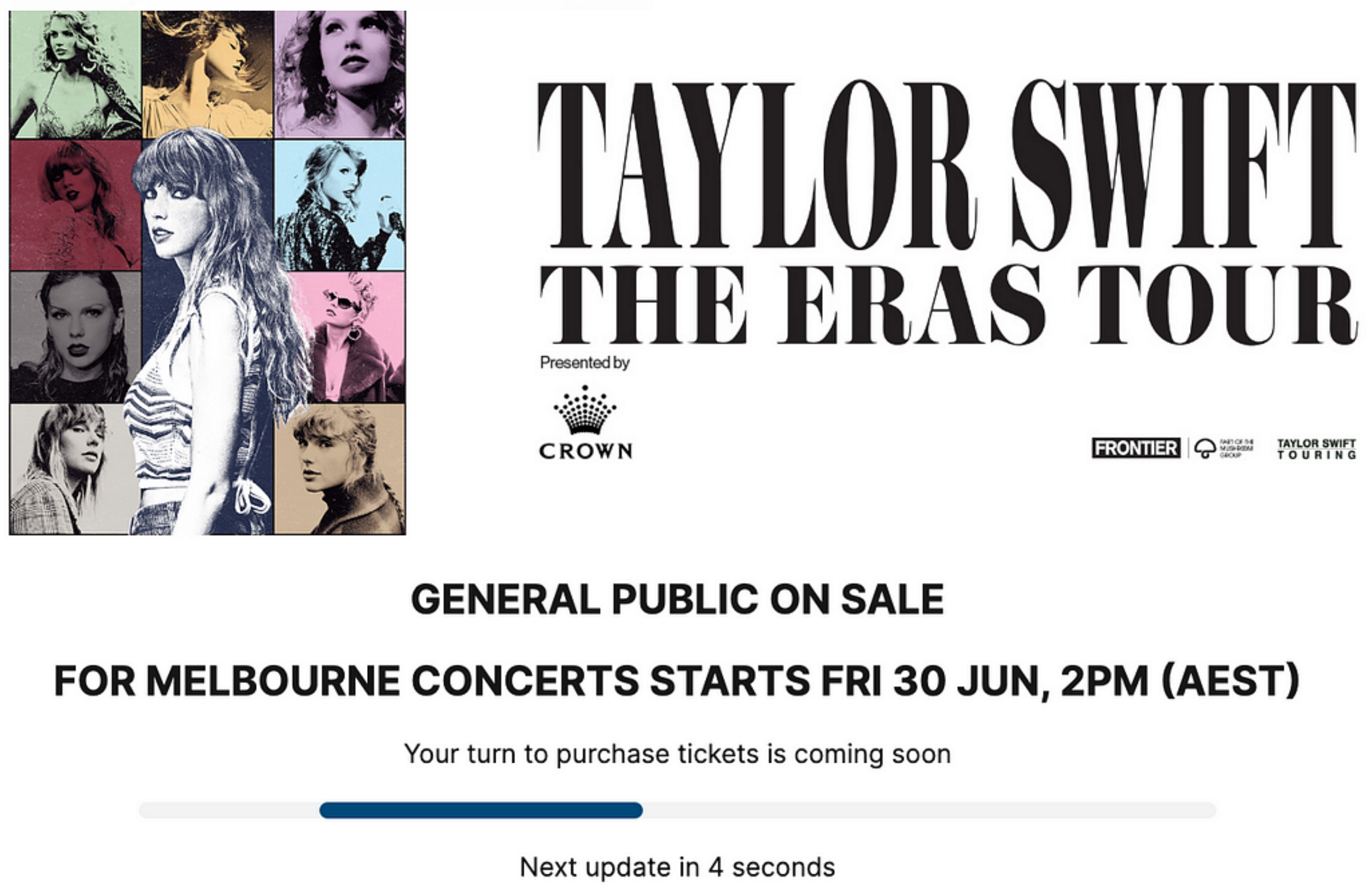
Transform Method for Markov-Modulated Queues

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

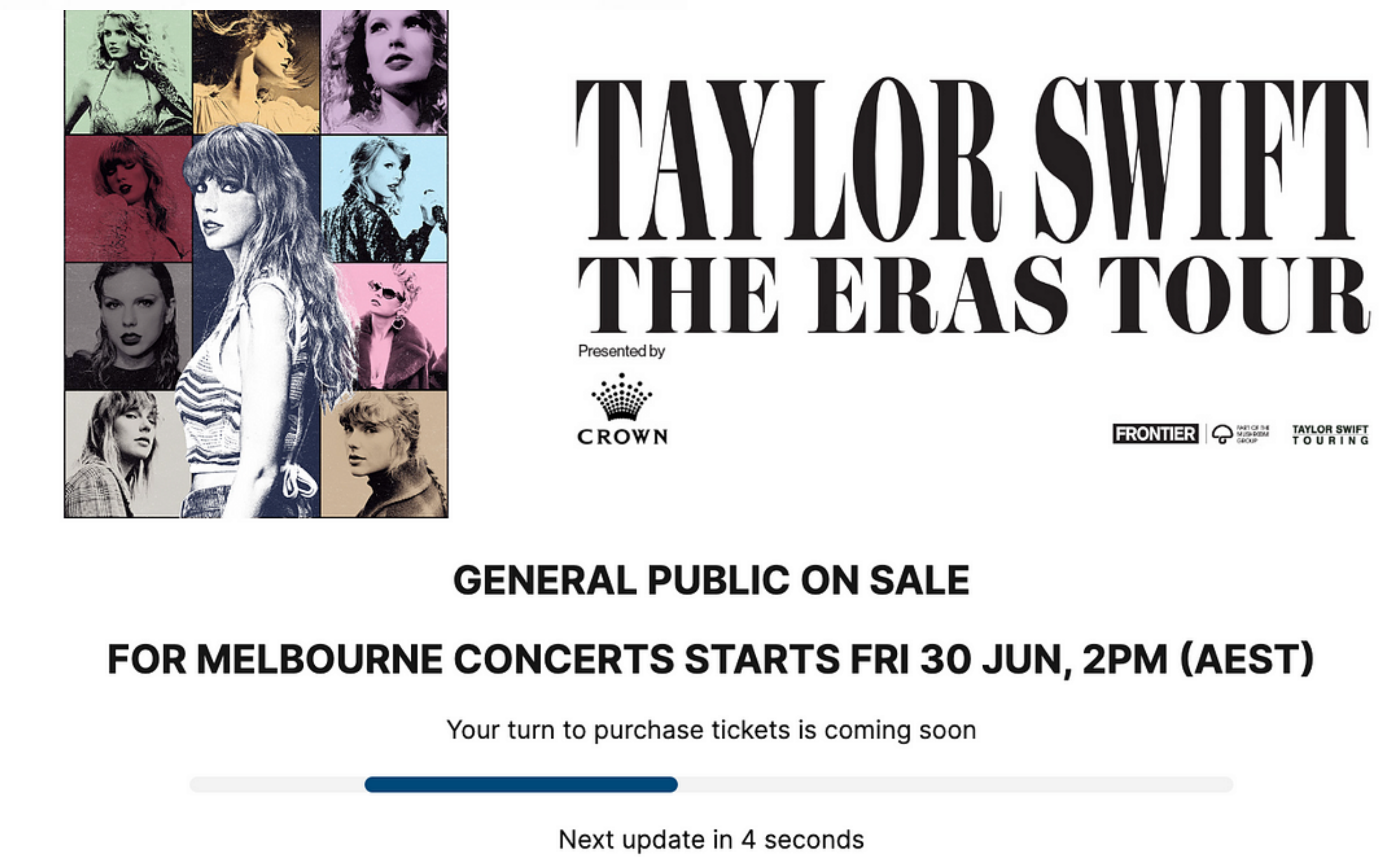
YEQT 2025 — November 3-5th, 2025

Waiting in Line

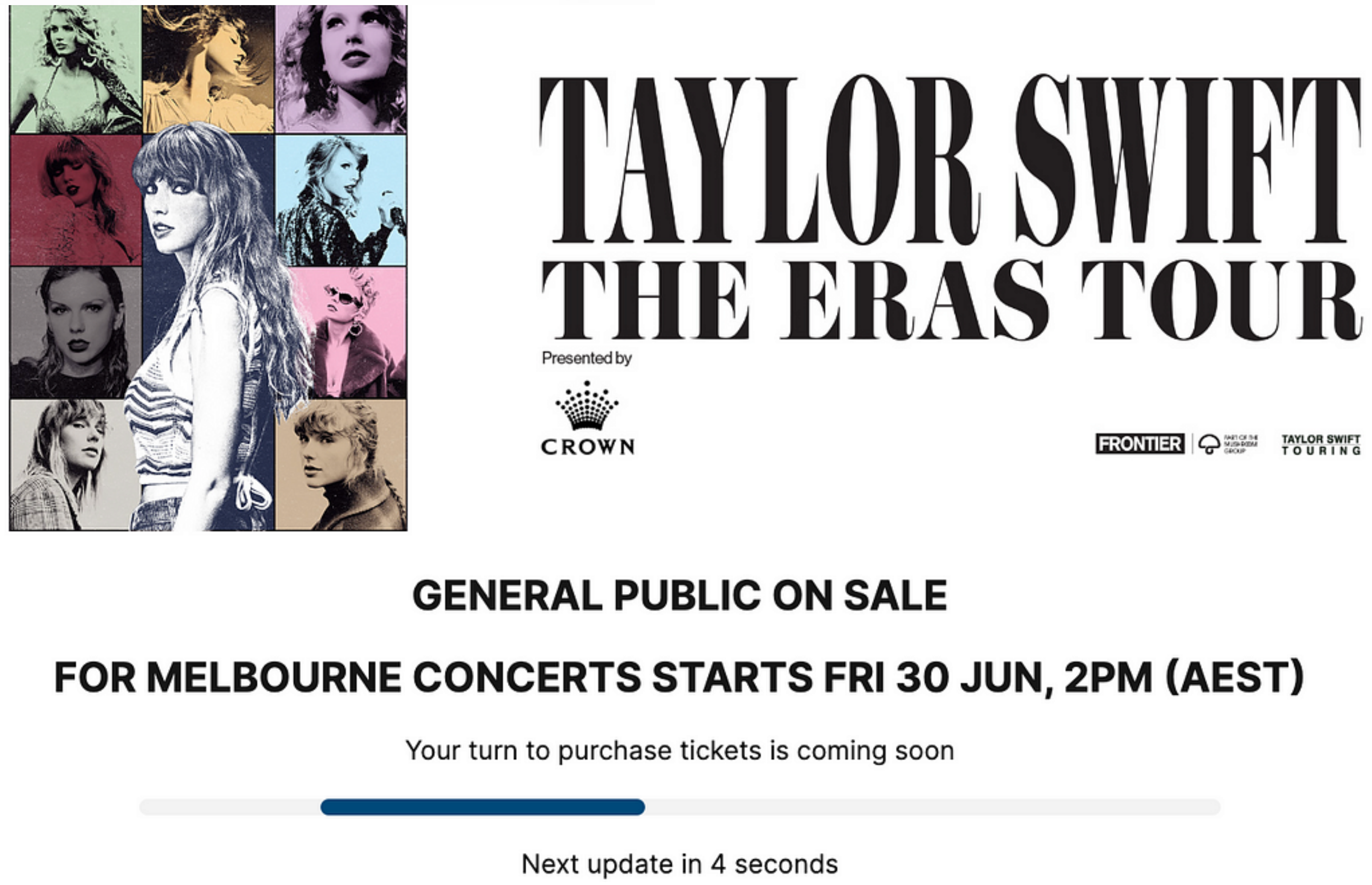
Waiting in Line

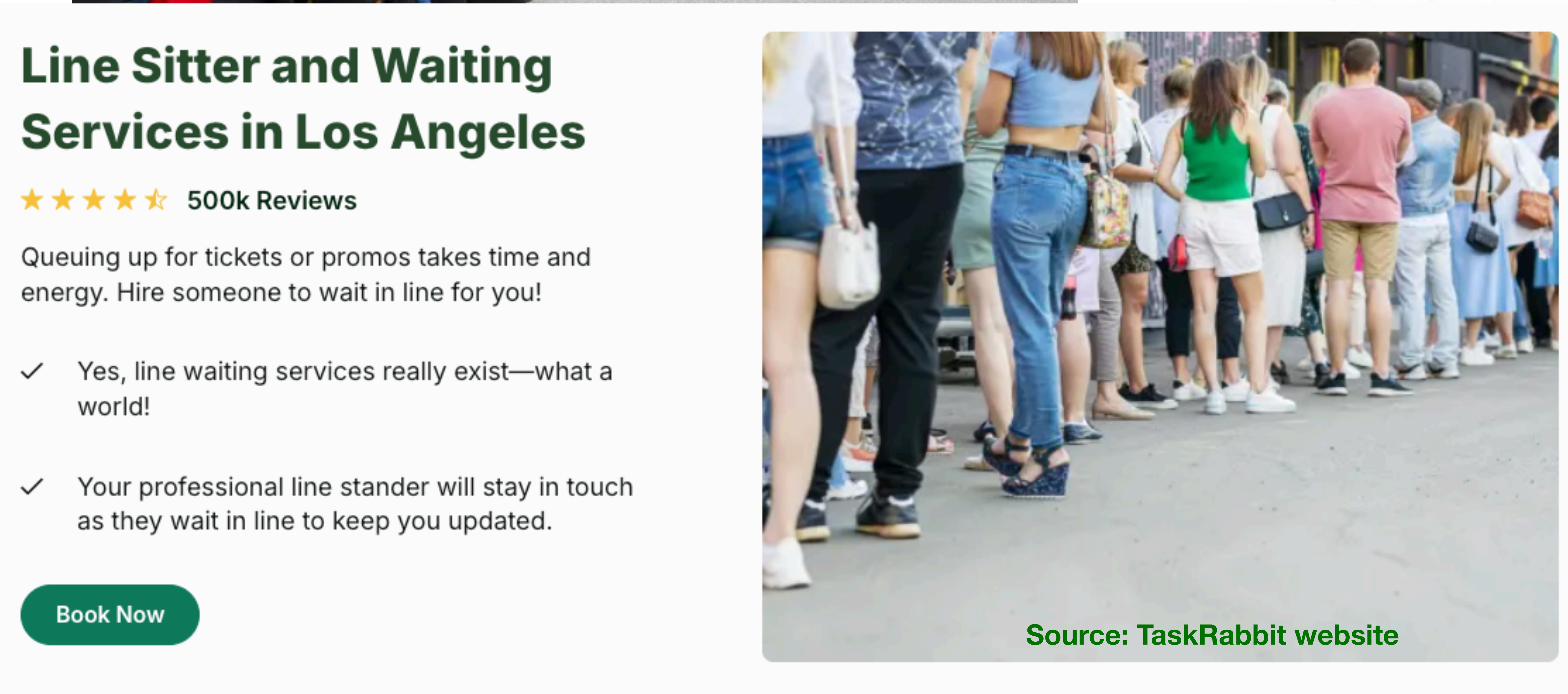


Waiting in Line



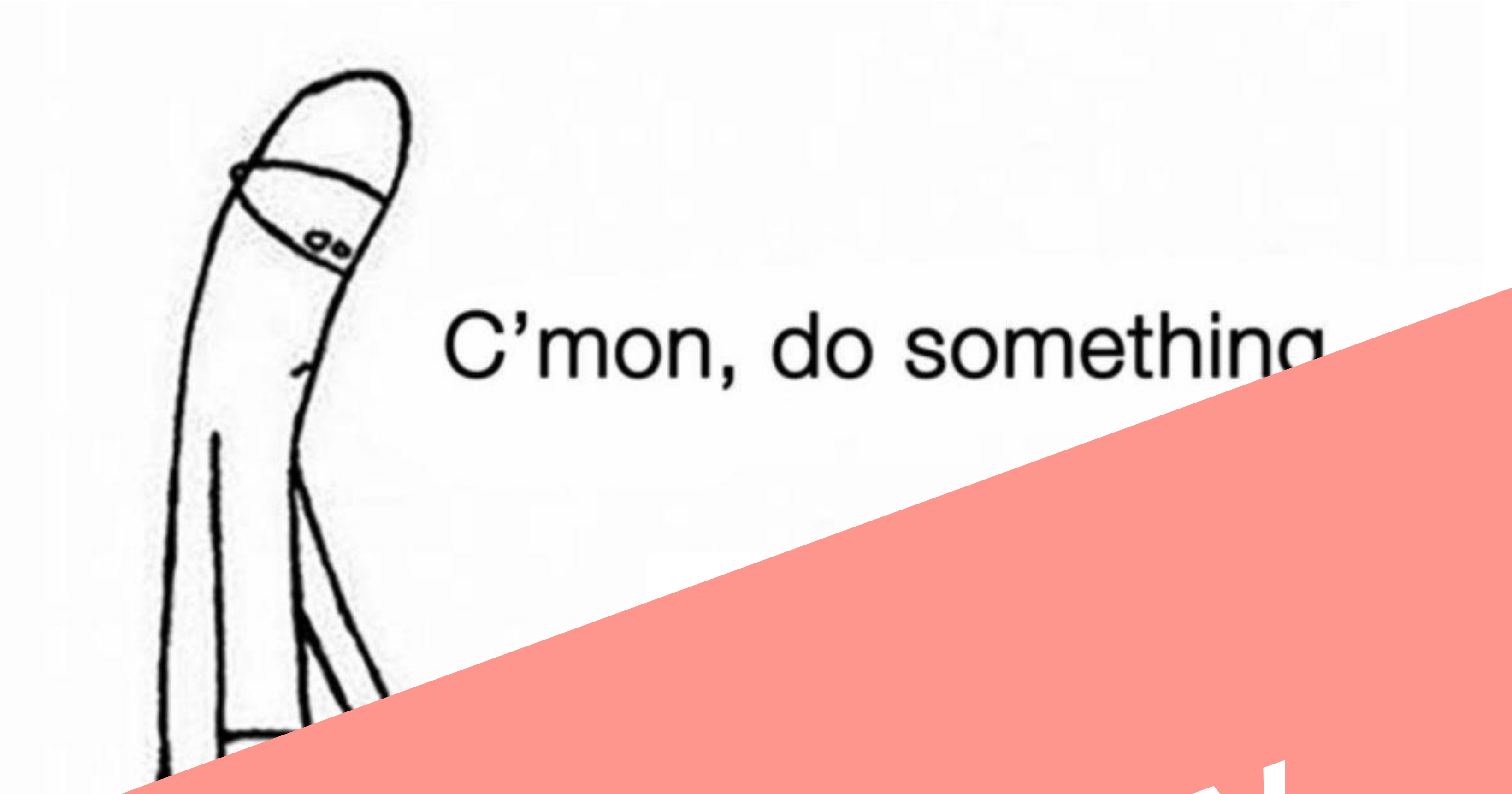
Waiting in Line

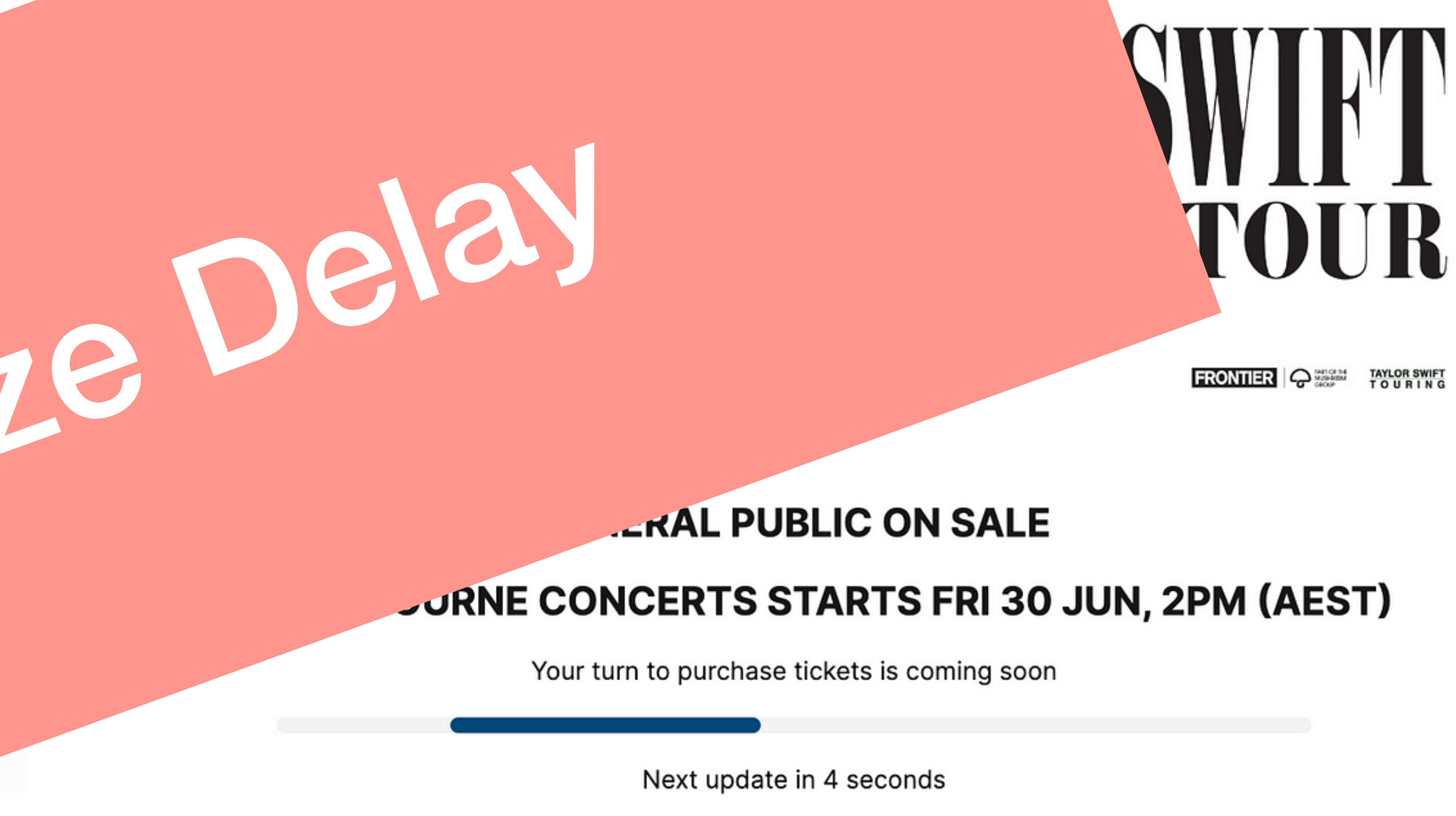




Source: TaskRabbit website

Waiting in Line



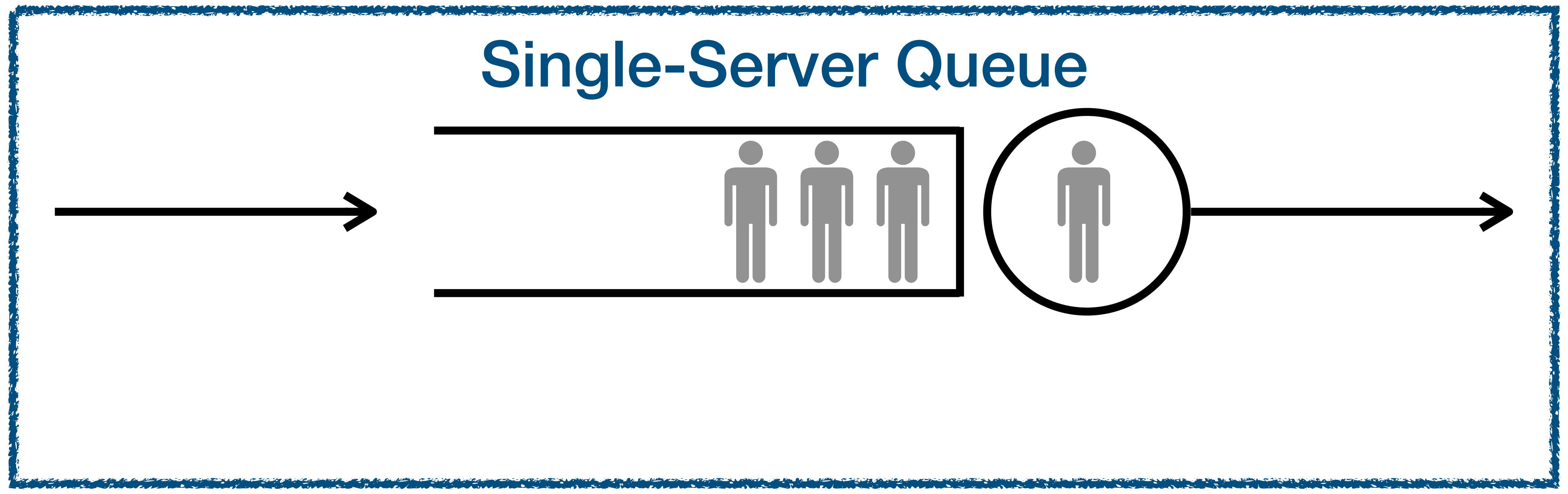


Source: TaskRabbit website

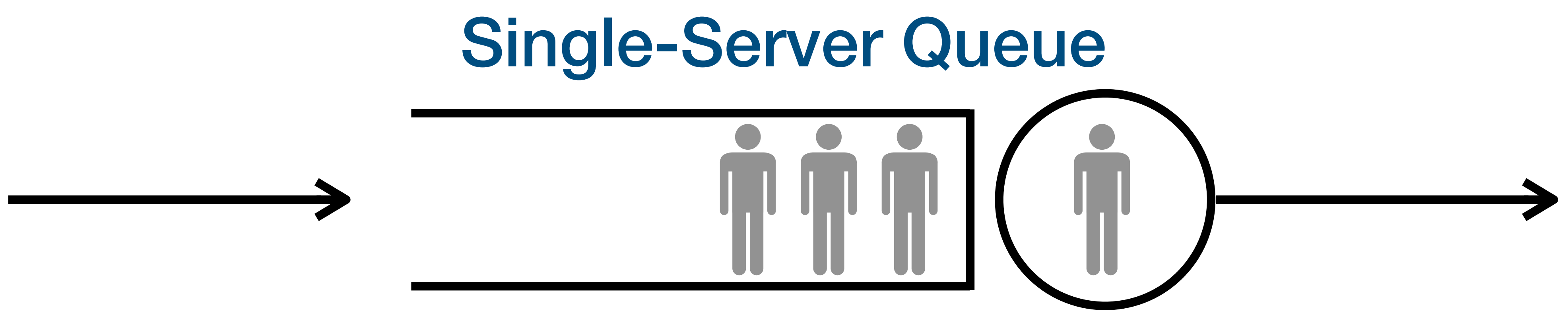
Ultimate goal: Minimize Delay

Understanding Delay

Understanding Delay



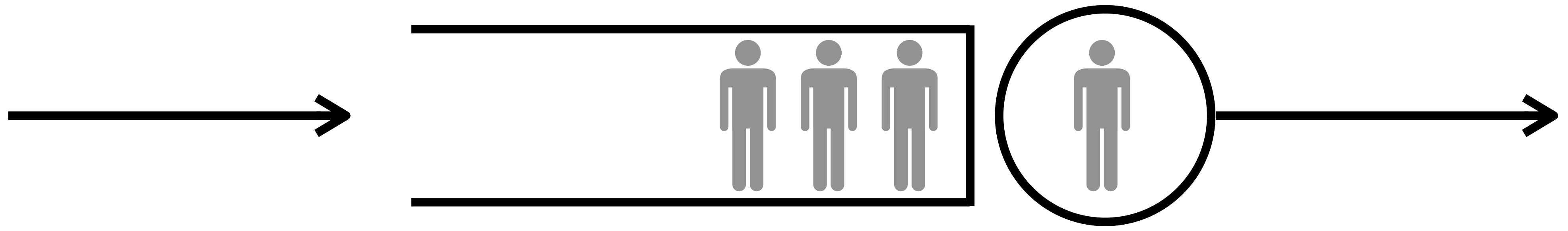
Understanding Delay



✓ Well understood
Distribution of q is known

Understanding Delay

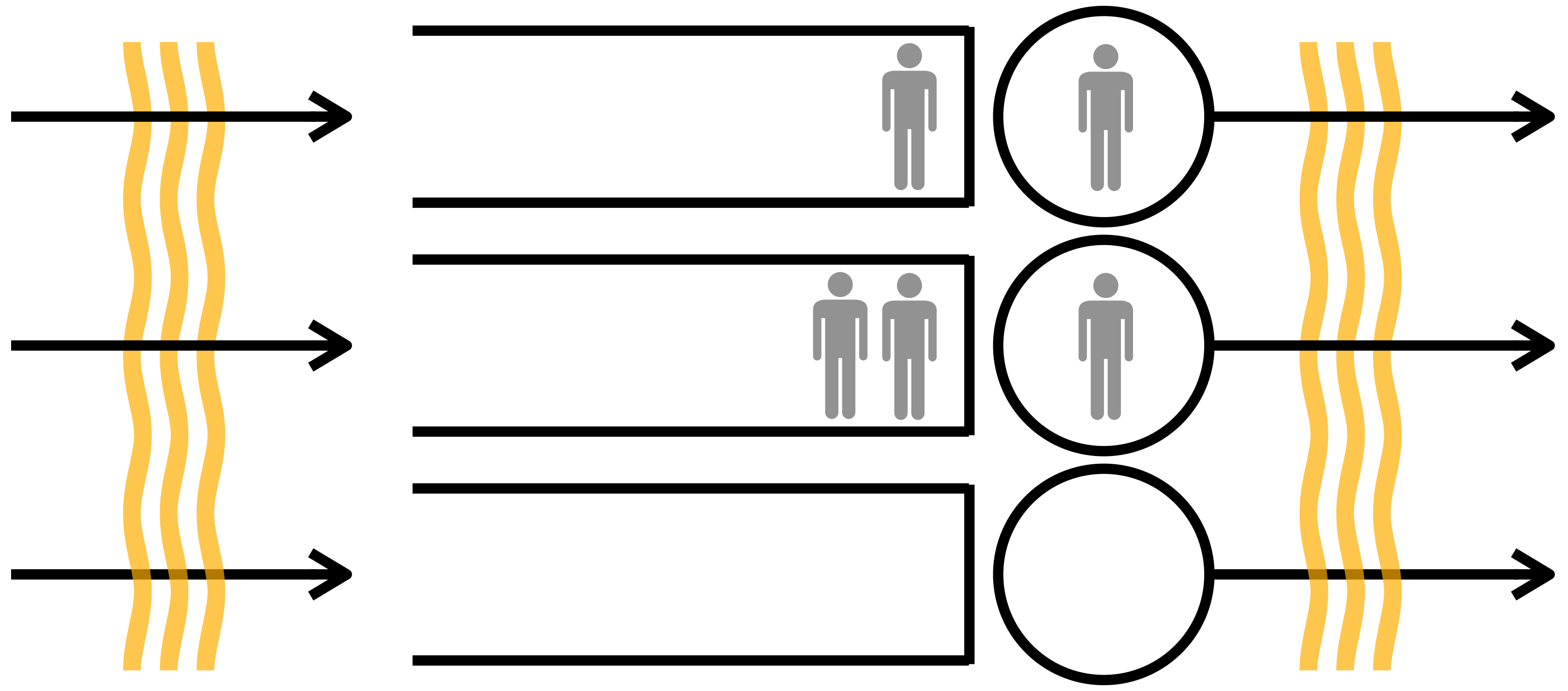
Single-Server Queue



✓ Well understood

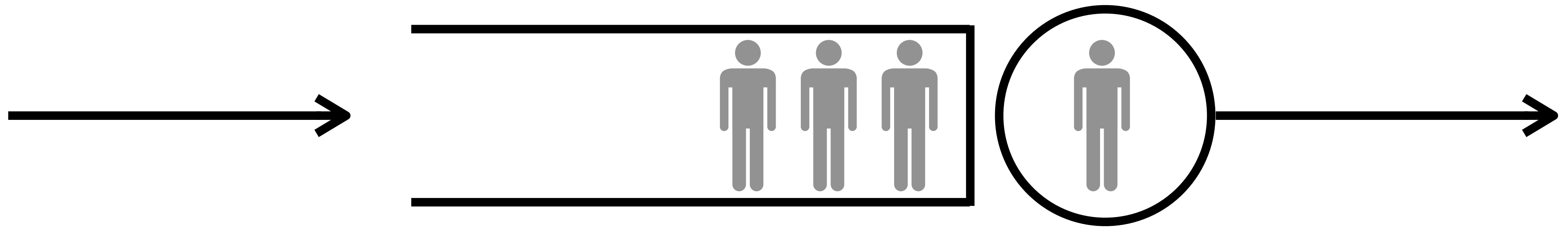
Distribution of q is known

Multiple queues with correlated arrivals and services



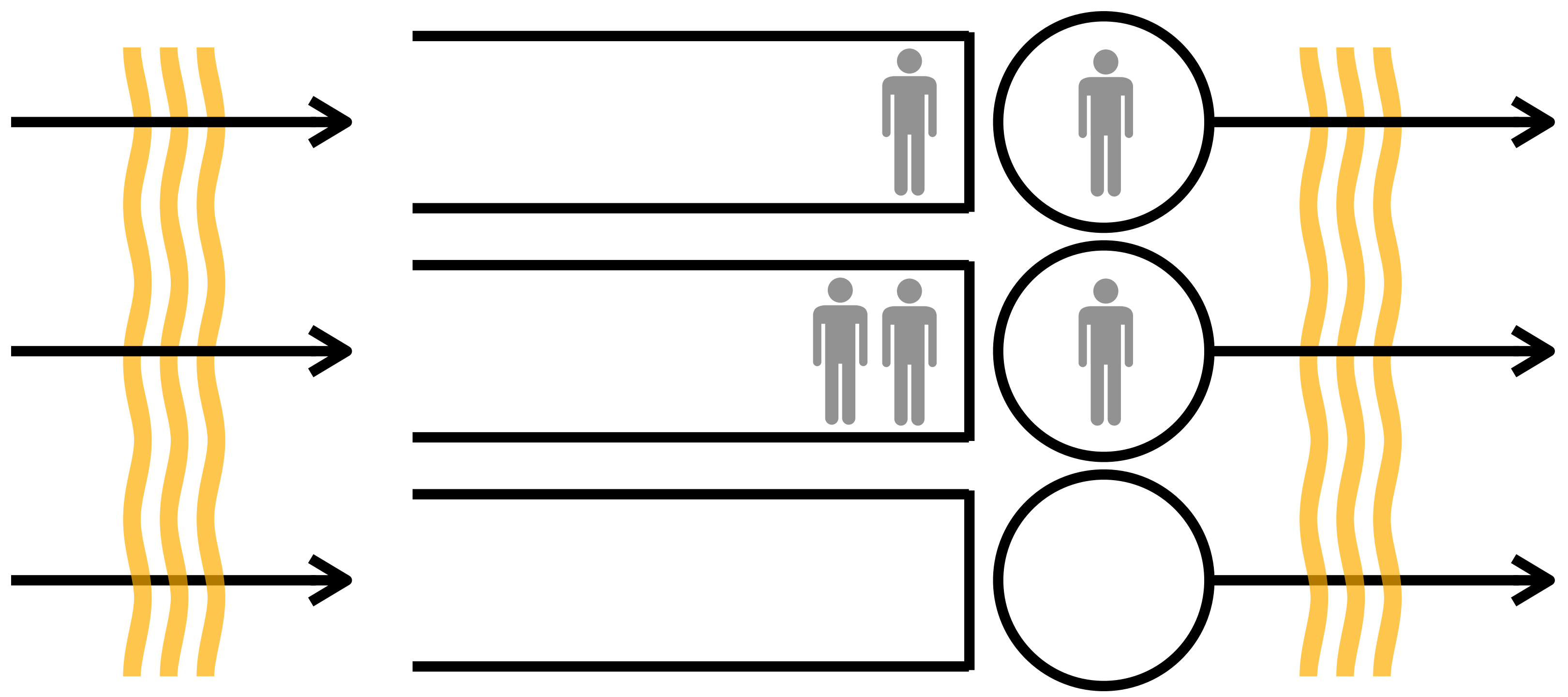
Understanding Delay

Single-Server Queue

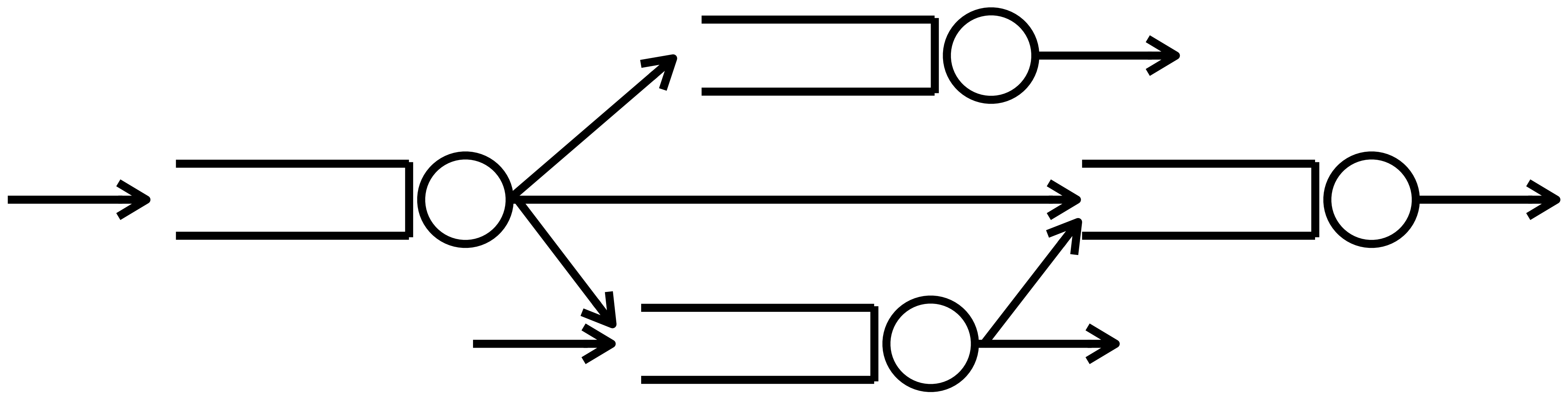


✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services

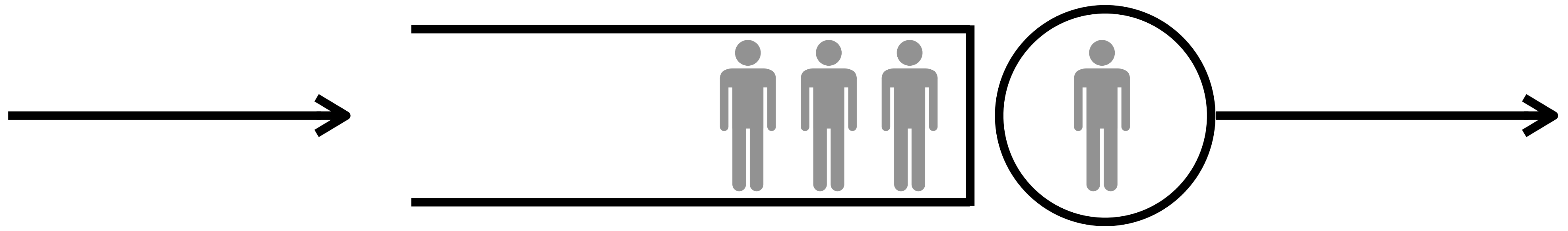


Networks of queues



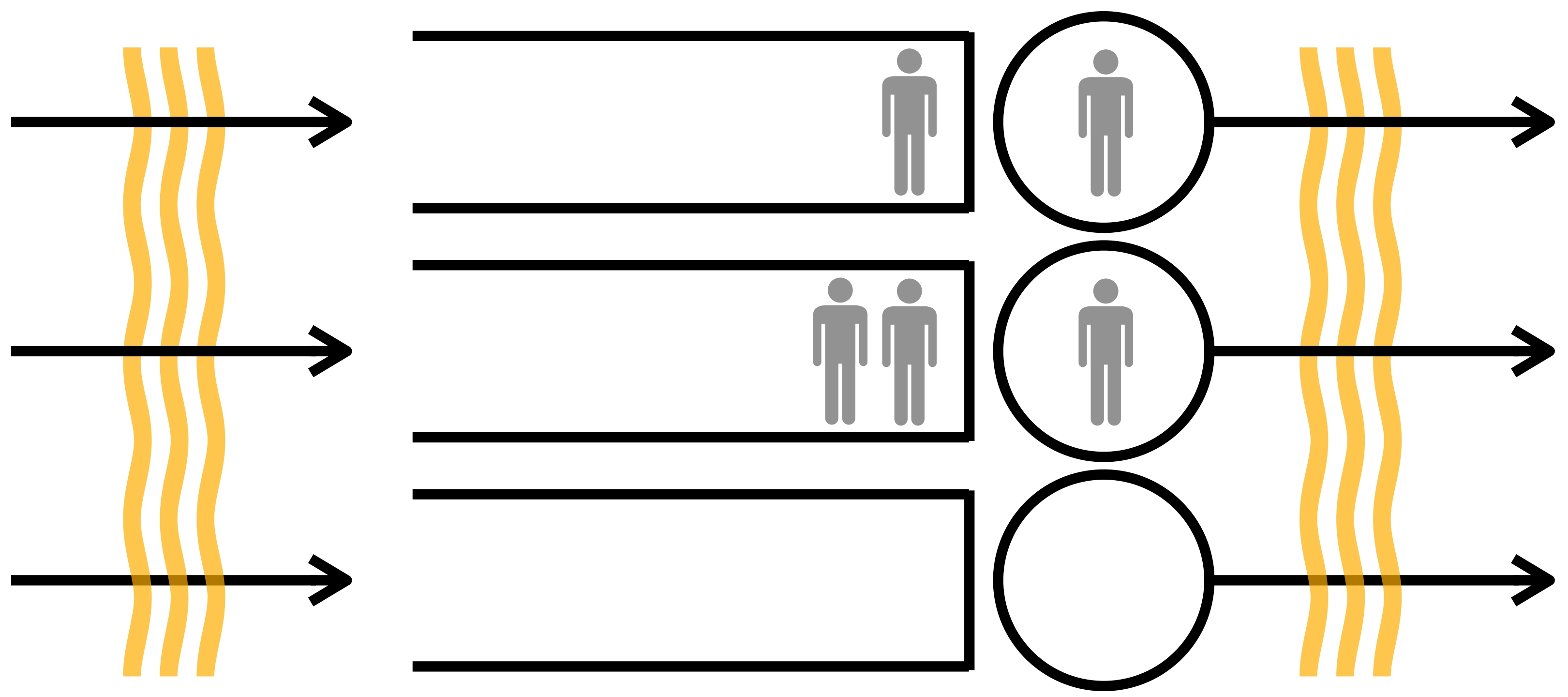
Understanding Delay

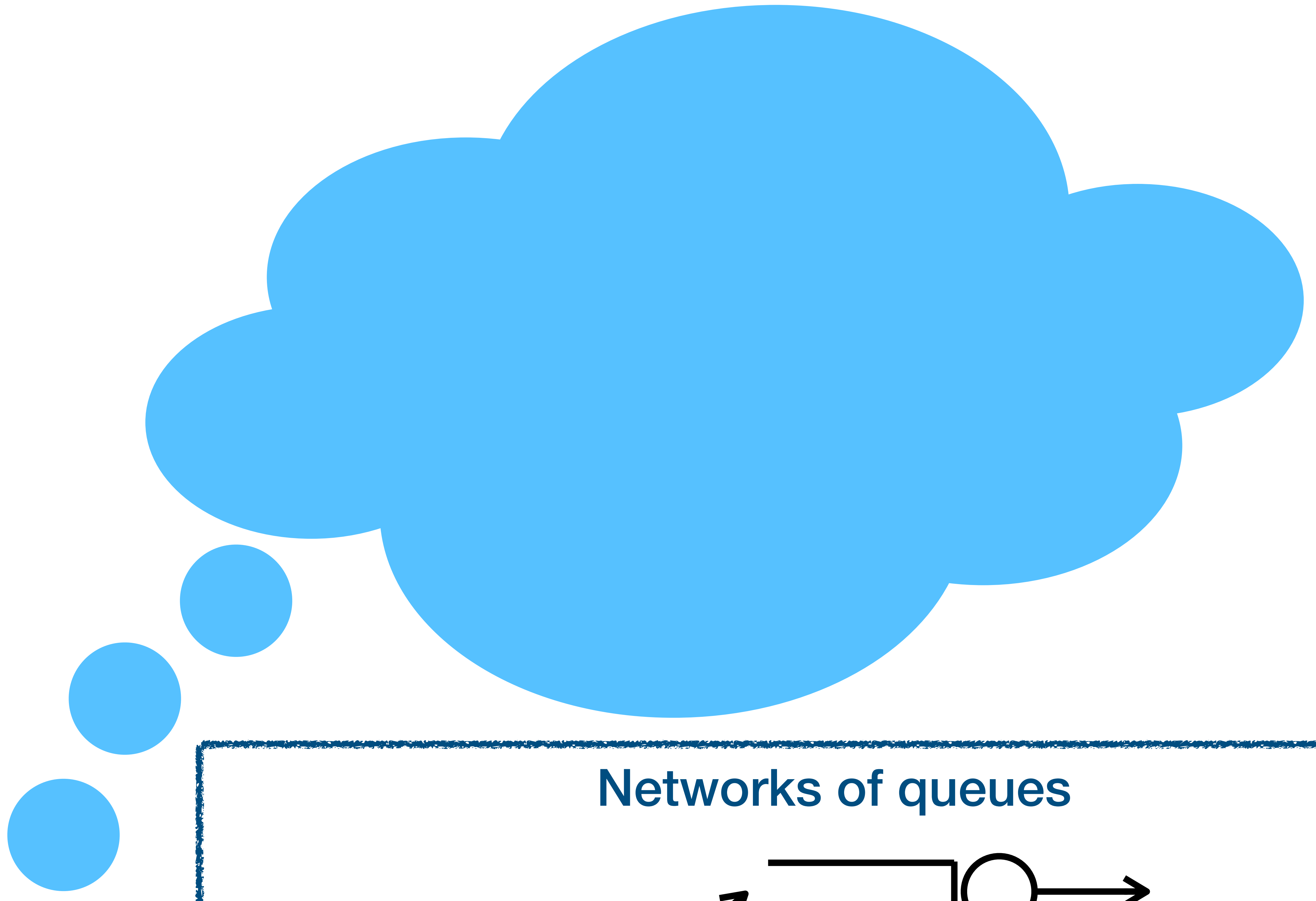
Single-Server Queue



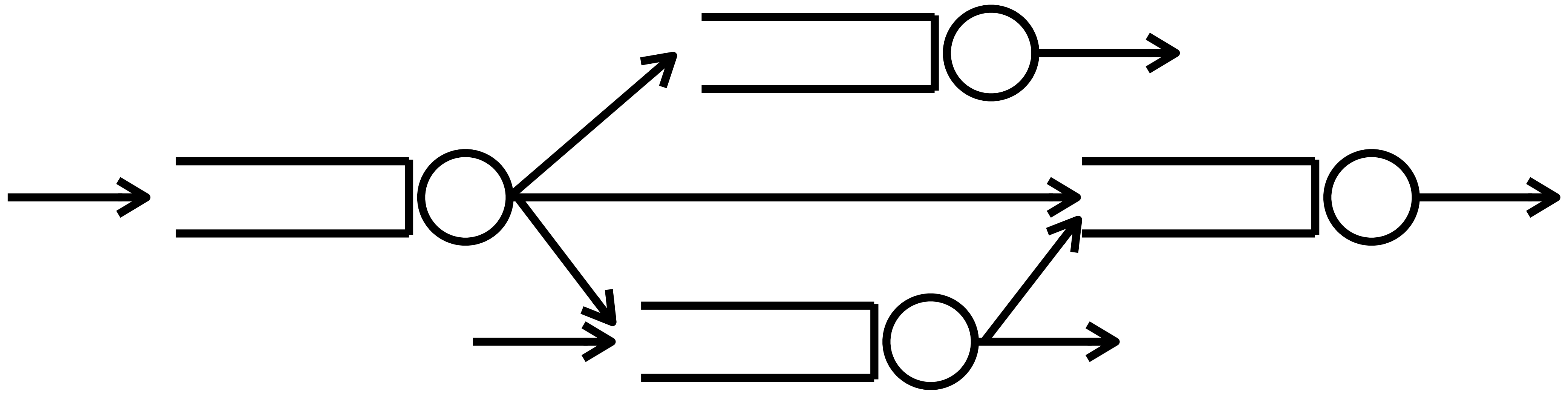
✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services





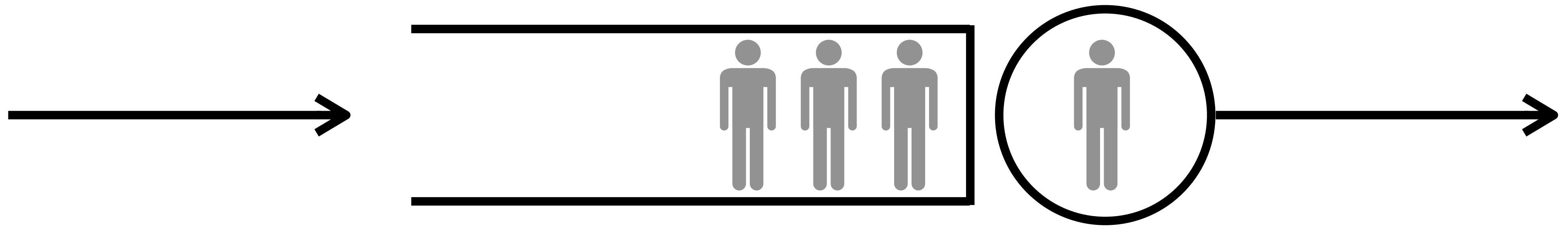
Networks of queues



Understanding Delay

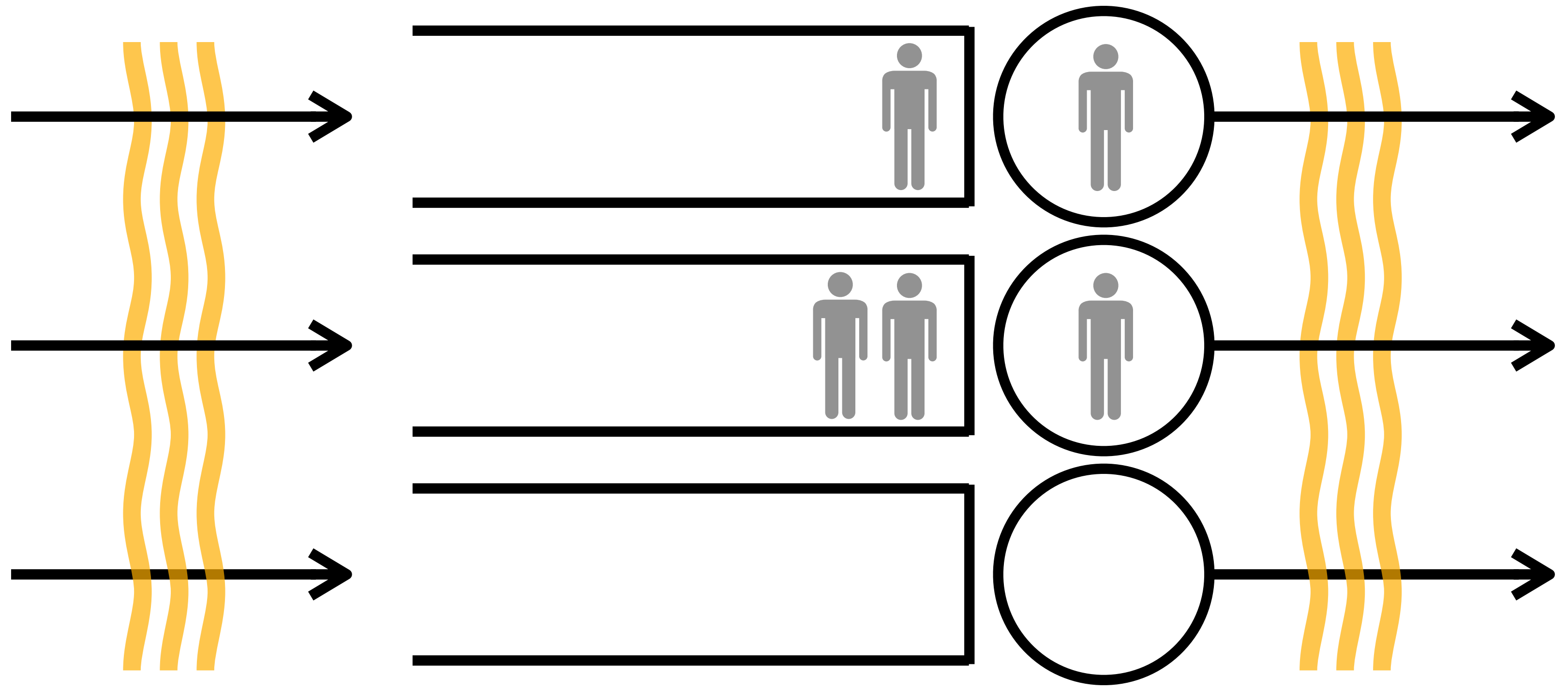
Distribution of q?

Single-Server Queue

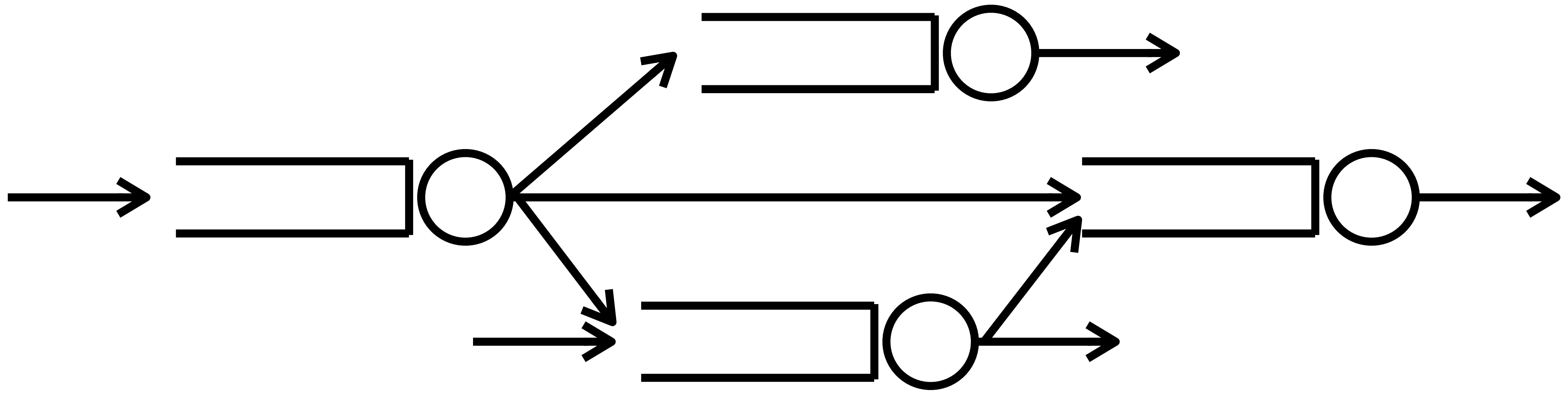


✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services



Networks of queues

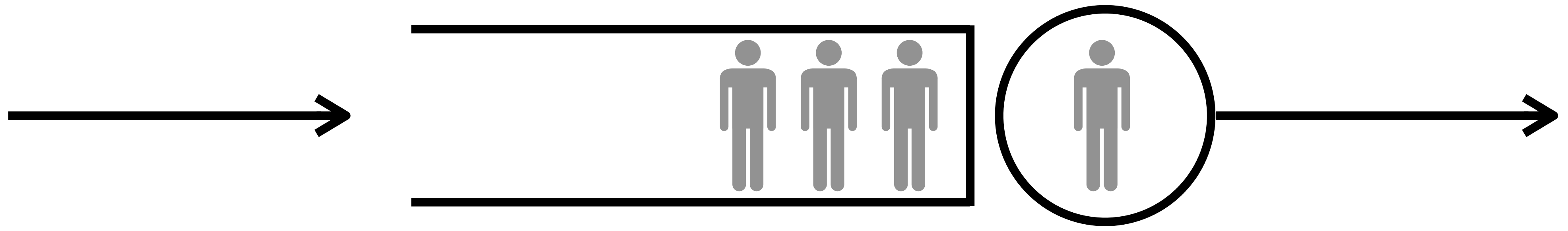


Understanding Delay

Distribution of q?

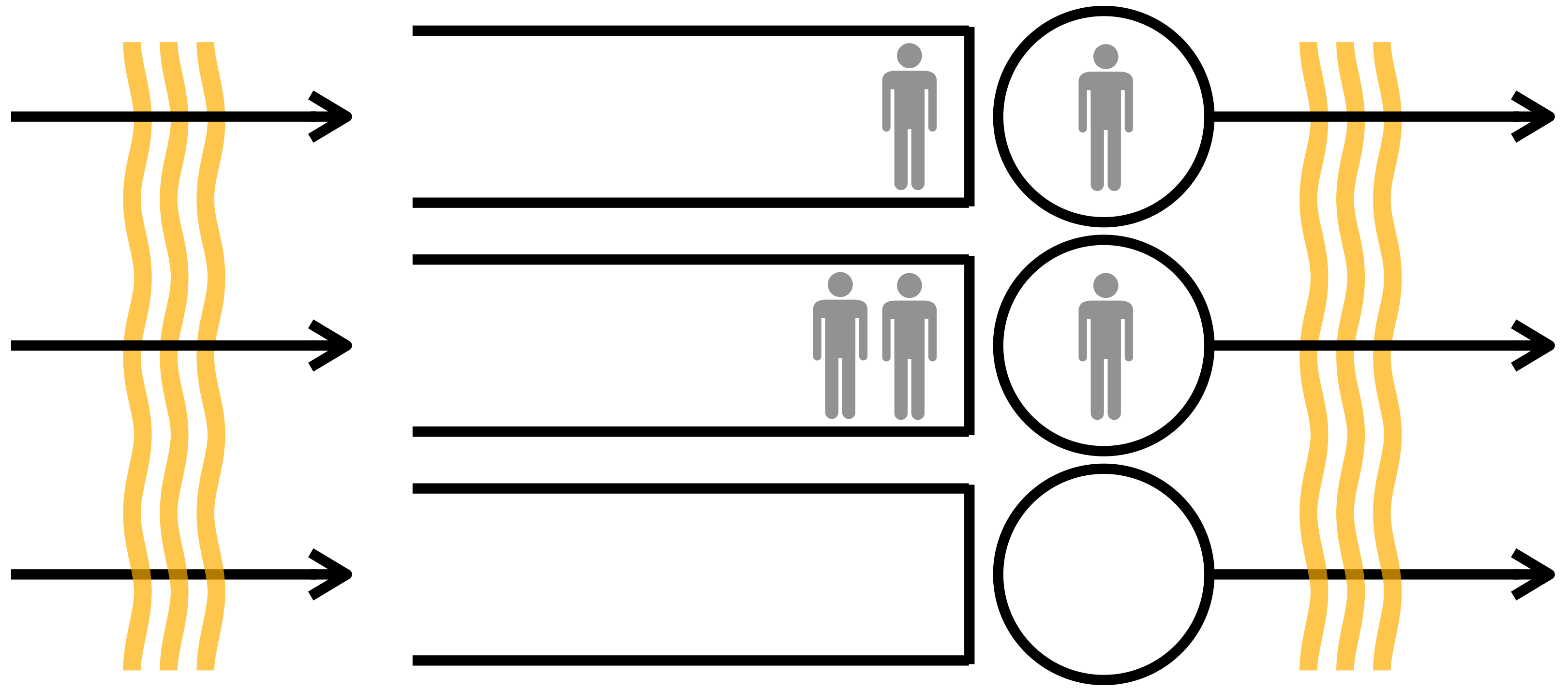
Mean: $E[q]$?

Single-Server Queue

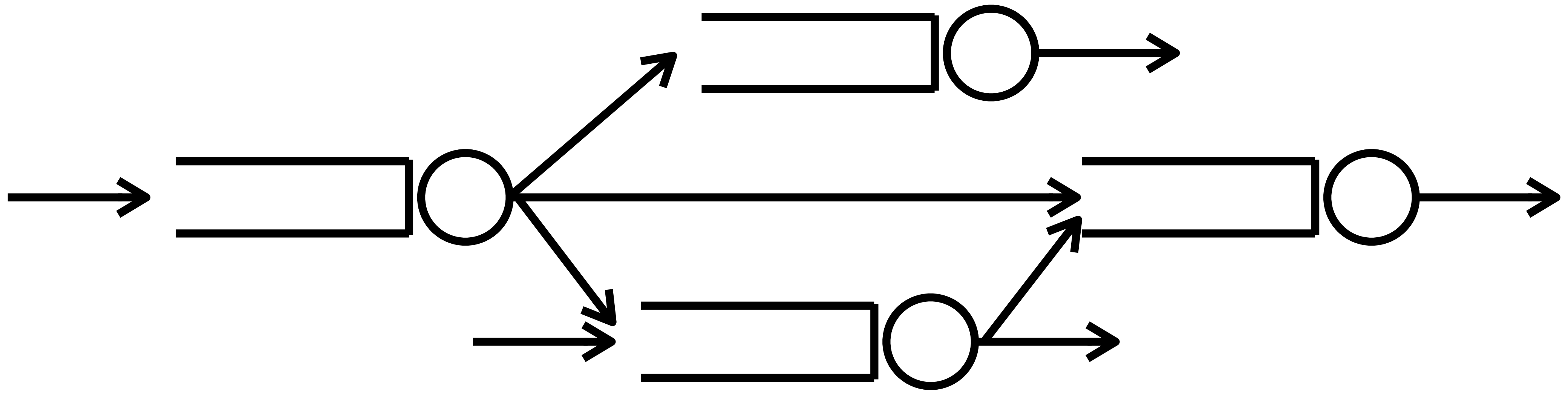


✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services

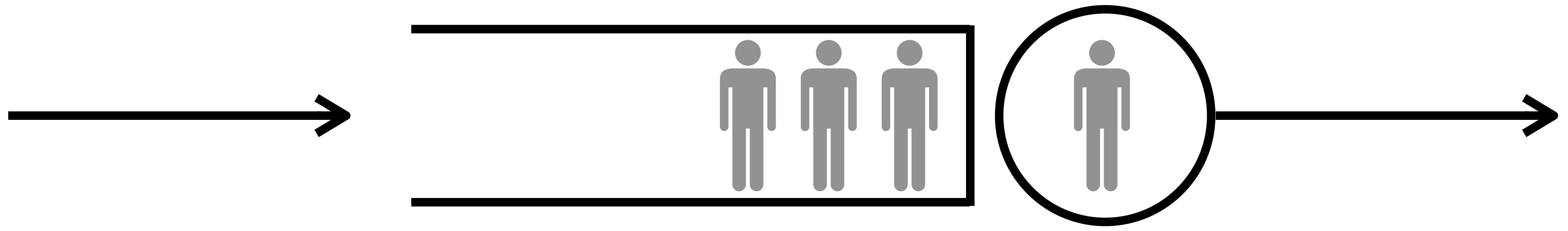


Networks of queues



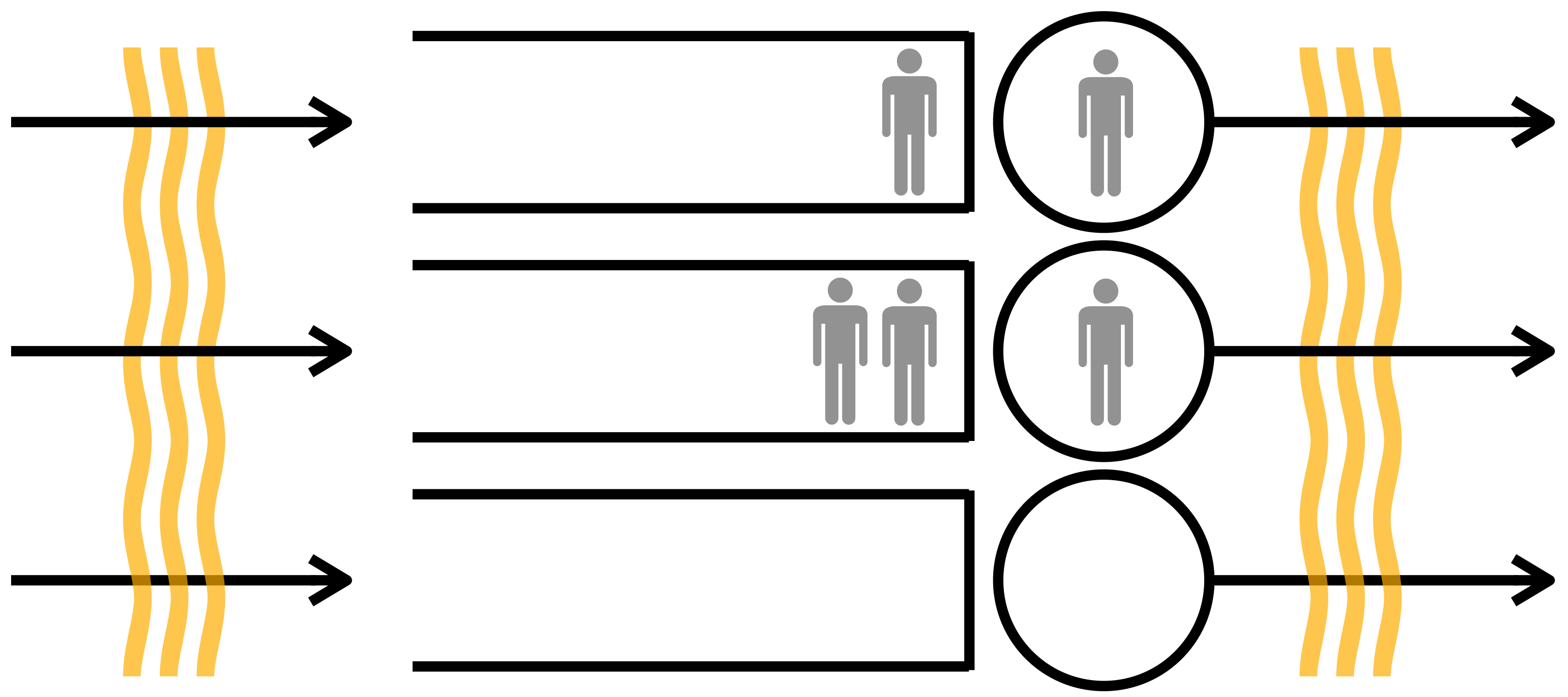
Understanding Delay

Single-Server Queue



✓ Well understood
Distribution of q is known

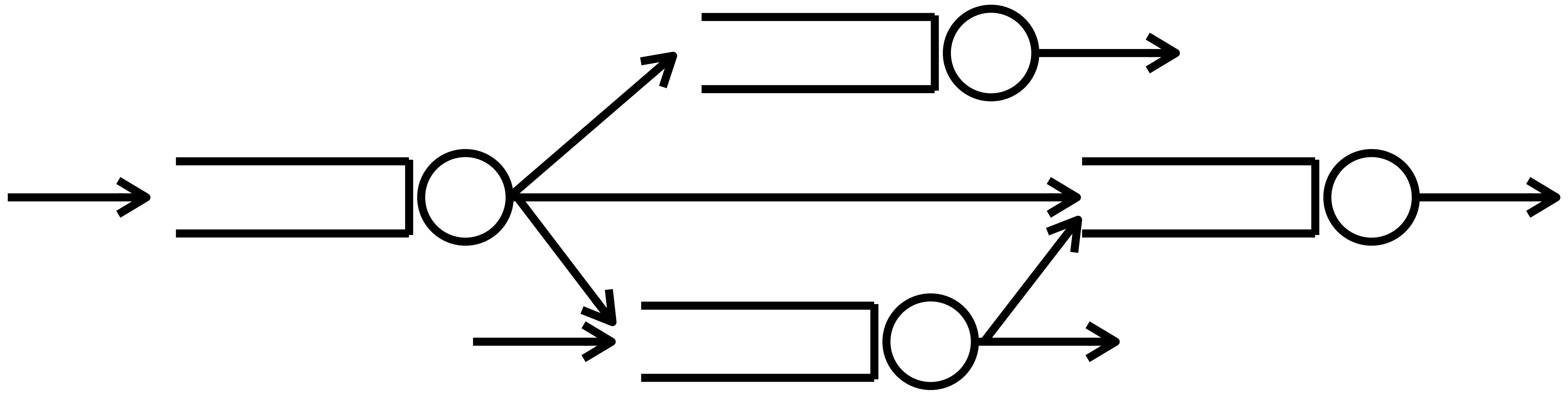
Multiple queues with correlated arrivals and services



Distribution of $\boldsymbol{q}$?

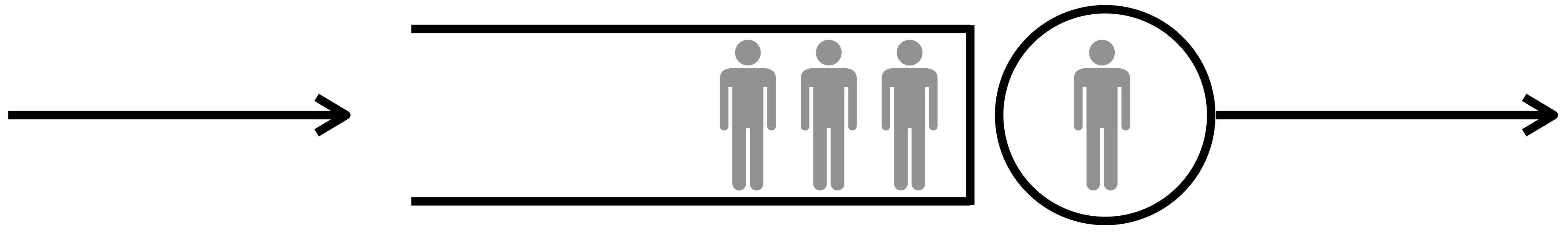
Mean: $E[\boldsymbol{q}]$? Variance: $Var[\boldsymbol{q}]$?

Networks of queues



Understanding Delay

Single-Server Queue



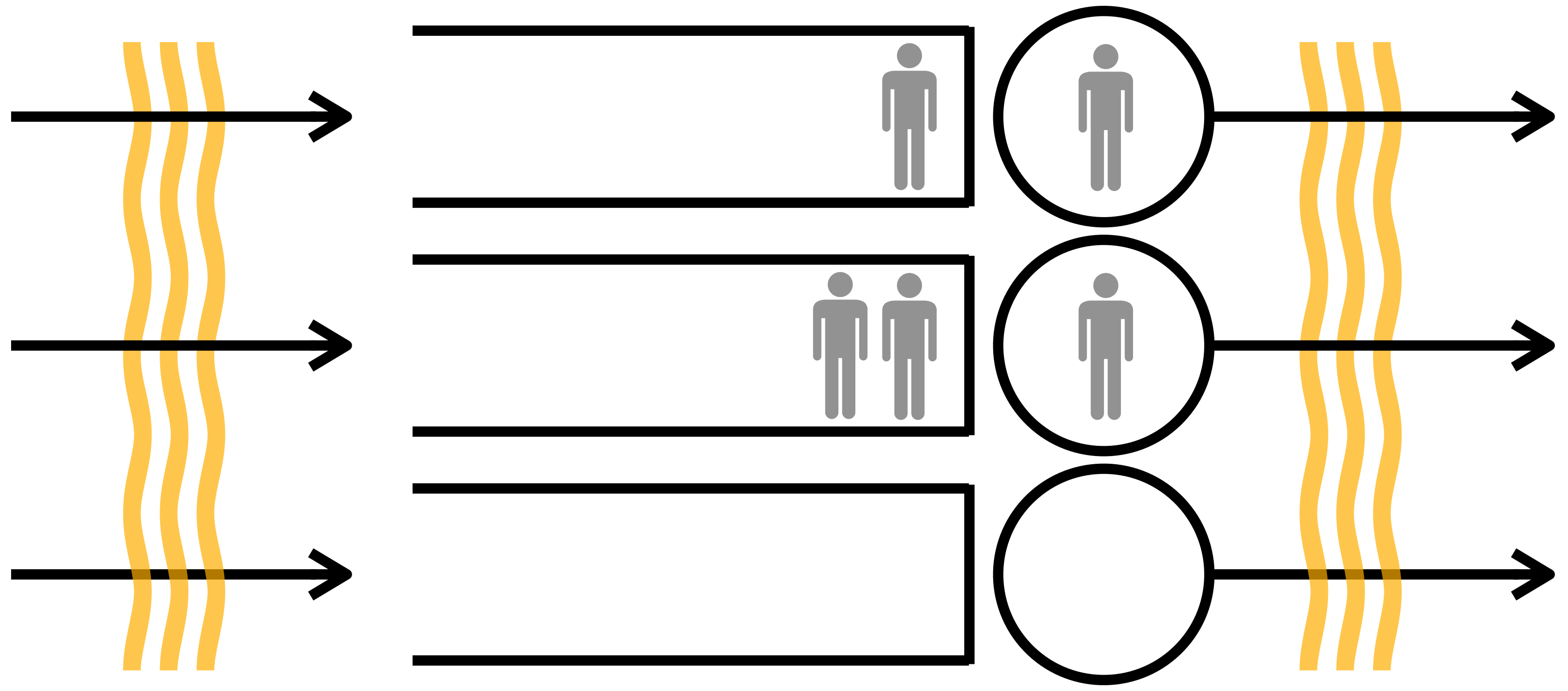
✓ Well understood
Distribution of q is known

Distribution of $\boldsymbol{q}$?

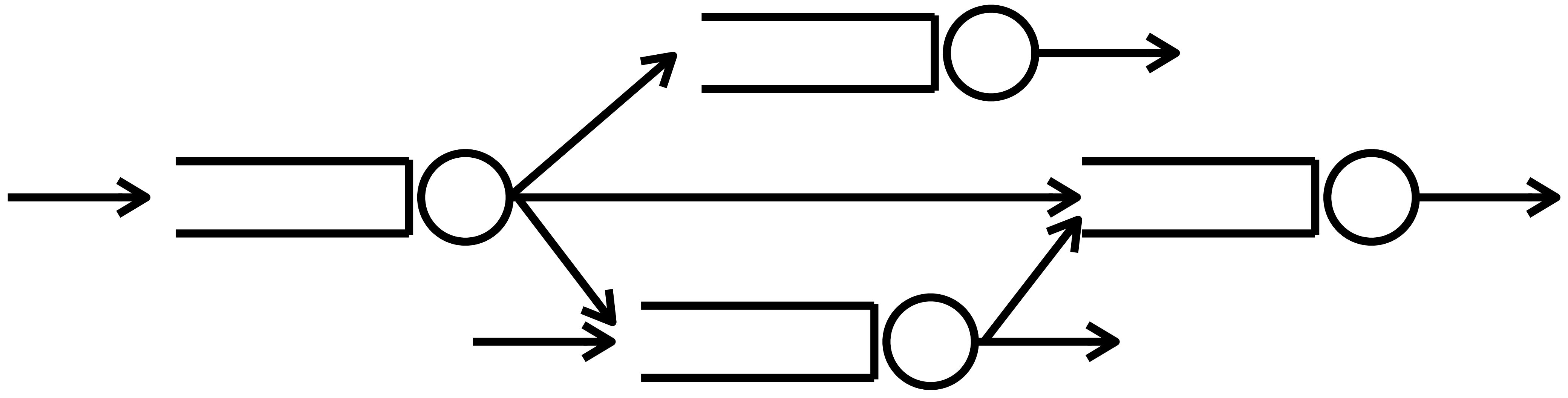
Mean: $E[\boldsymbol{q}]$?   Variance: $Var[\boldsymbol{q}]$?

How about $\sum q_j$?

Multiple queues with correlated arrivals and services

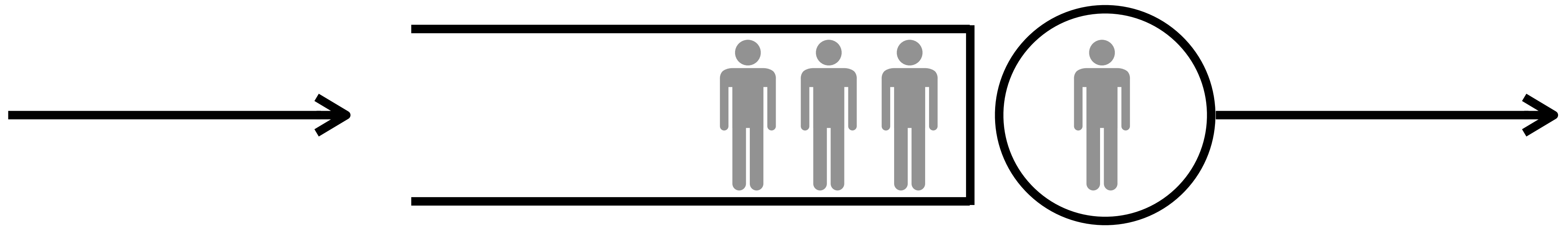


Networks of queues



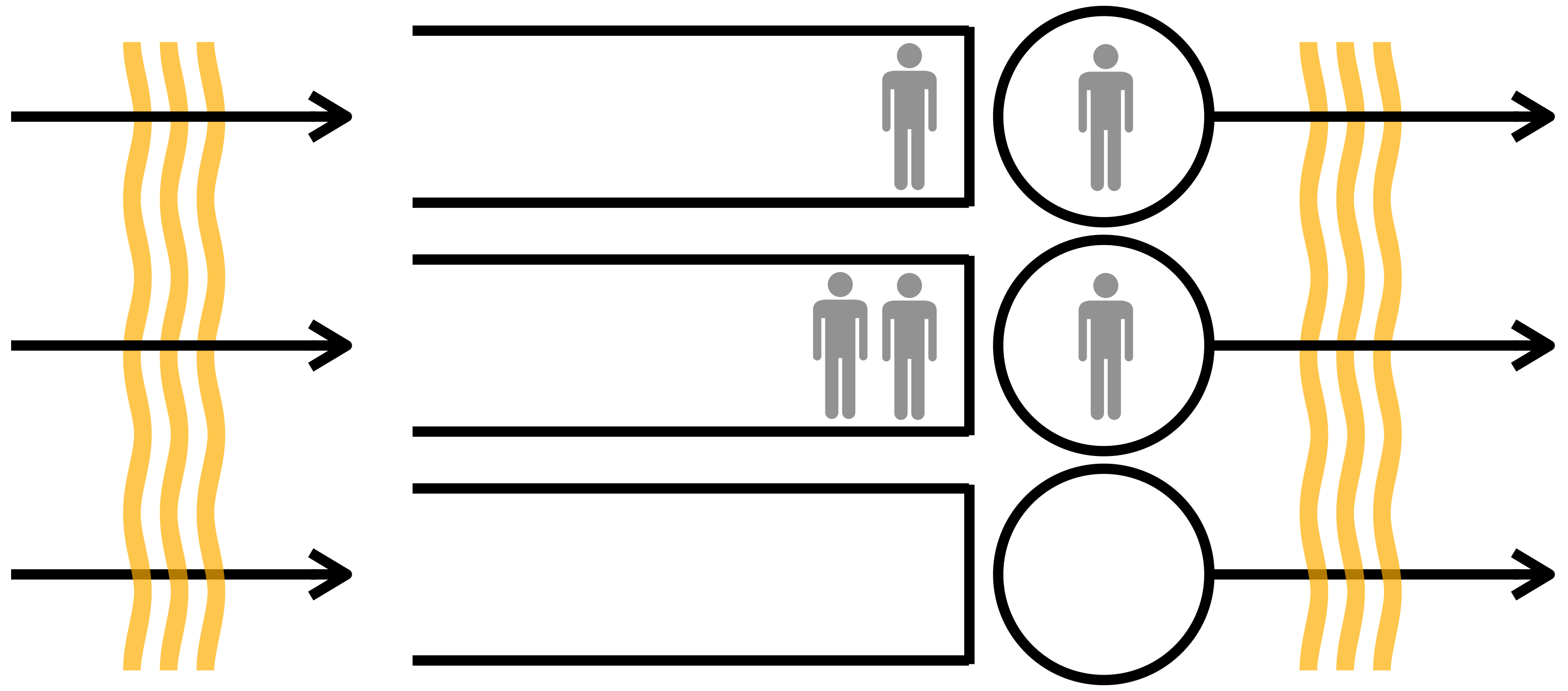
Understanding Delay

Single-Server Queue



✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services



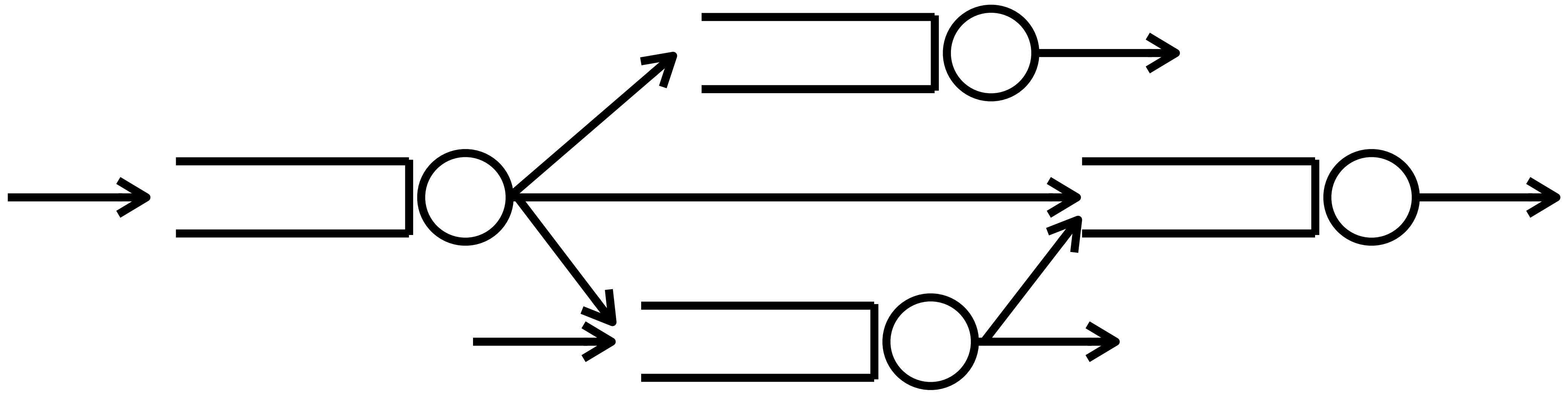
Distribution of $\boldsymbol{q}$?

Mean: $E[\boldsymbol{q}]$? Variance: $Var[\boldsymbol{q}]$?

How about $\sum q_j$?

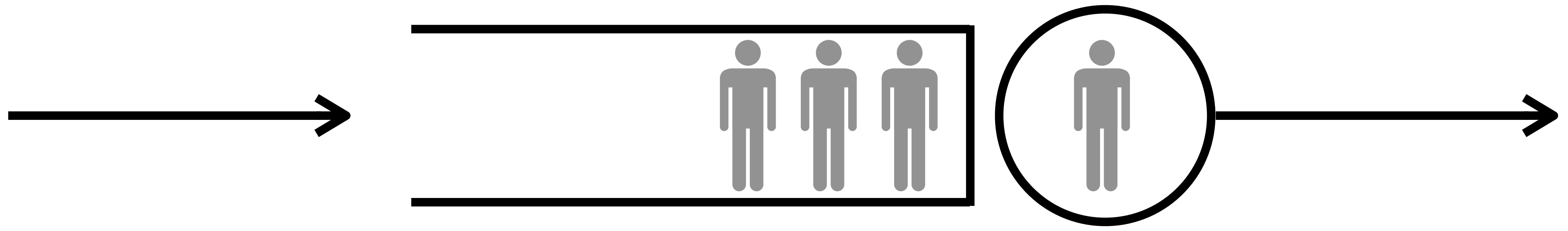
Tail prob: $P(\sum q_j > x)$?

Networks of queues



Understanding Delay

Single-Server Queue



✓ Well understood
Distribution of q is known

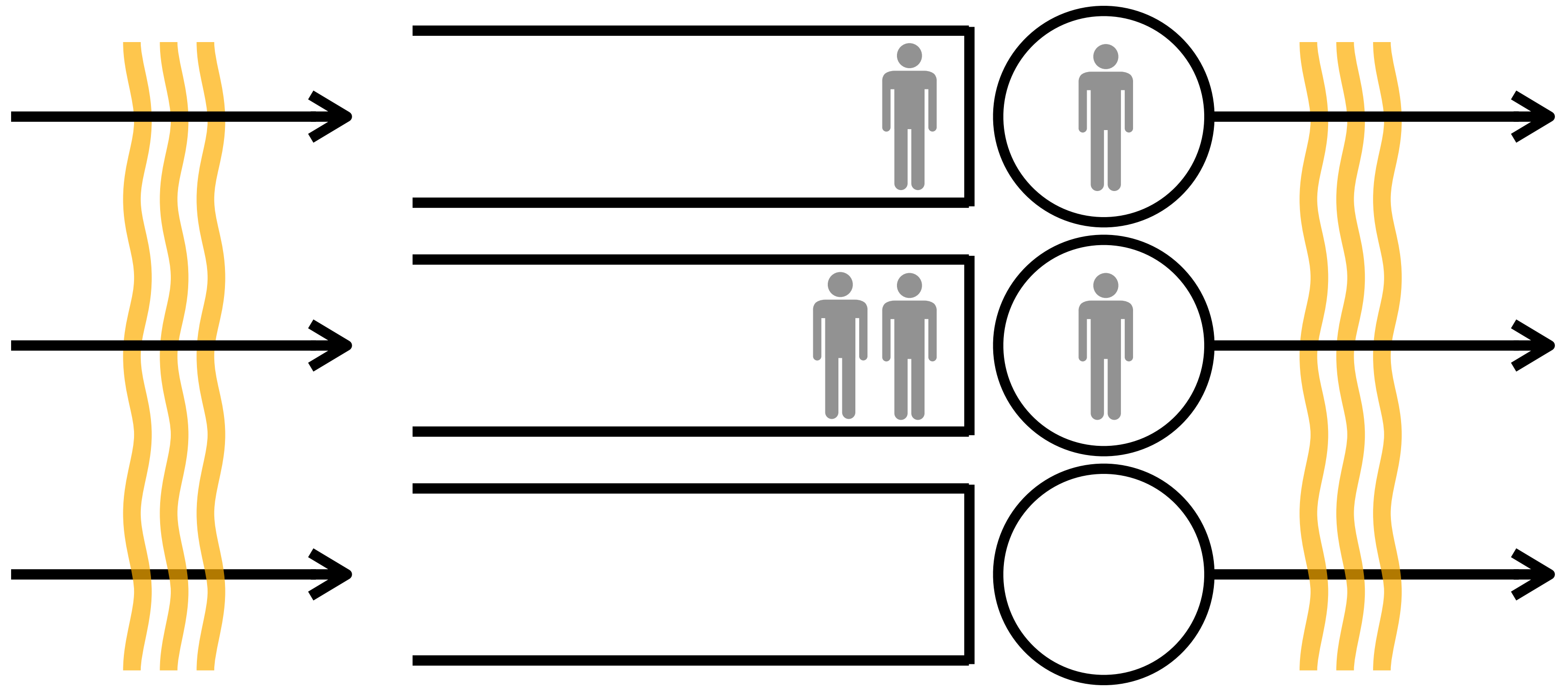
Distribution of $\boldsymbol{q}$?

Mean: $E[\boldsymbol{q}]$?   Variance: $Var[\boldsymbol{q}]$?

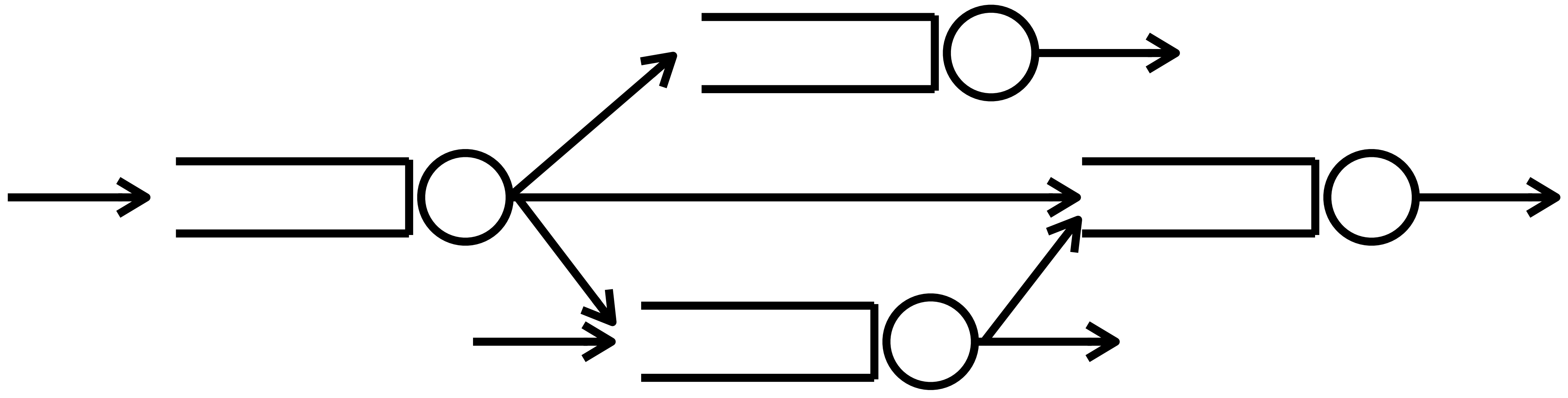
How about $\sum q_j$?

Tail prob: $P(\sum q_j > x)$?

Multiple queues with correlated arrivals and services



Networks of queues



✓ Jackson networks:
Distribution is product form

Understanding Delay

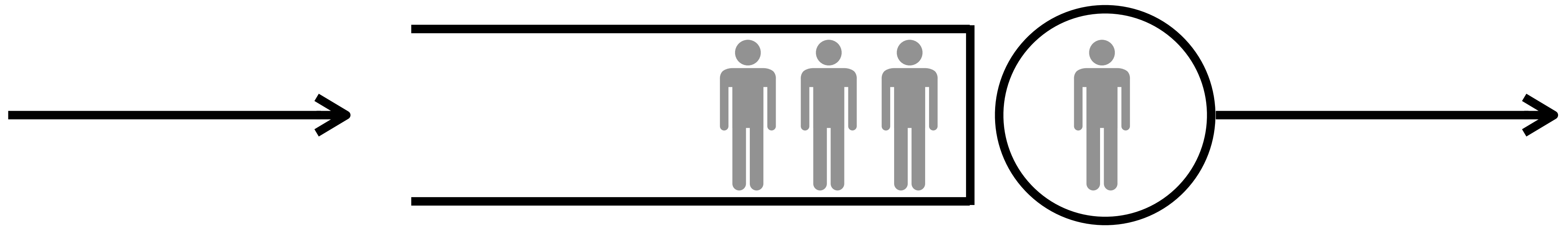
Distribution of $\boldsymbol{q}$?

Mean: $E[\boldsymbol{q}]$? Variance: $Var[\boldsymbol{q}]$?

How about $\sum q_j$?

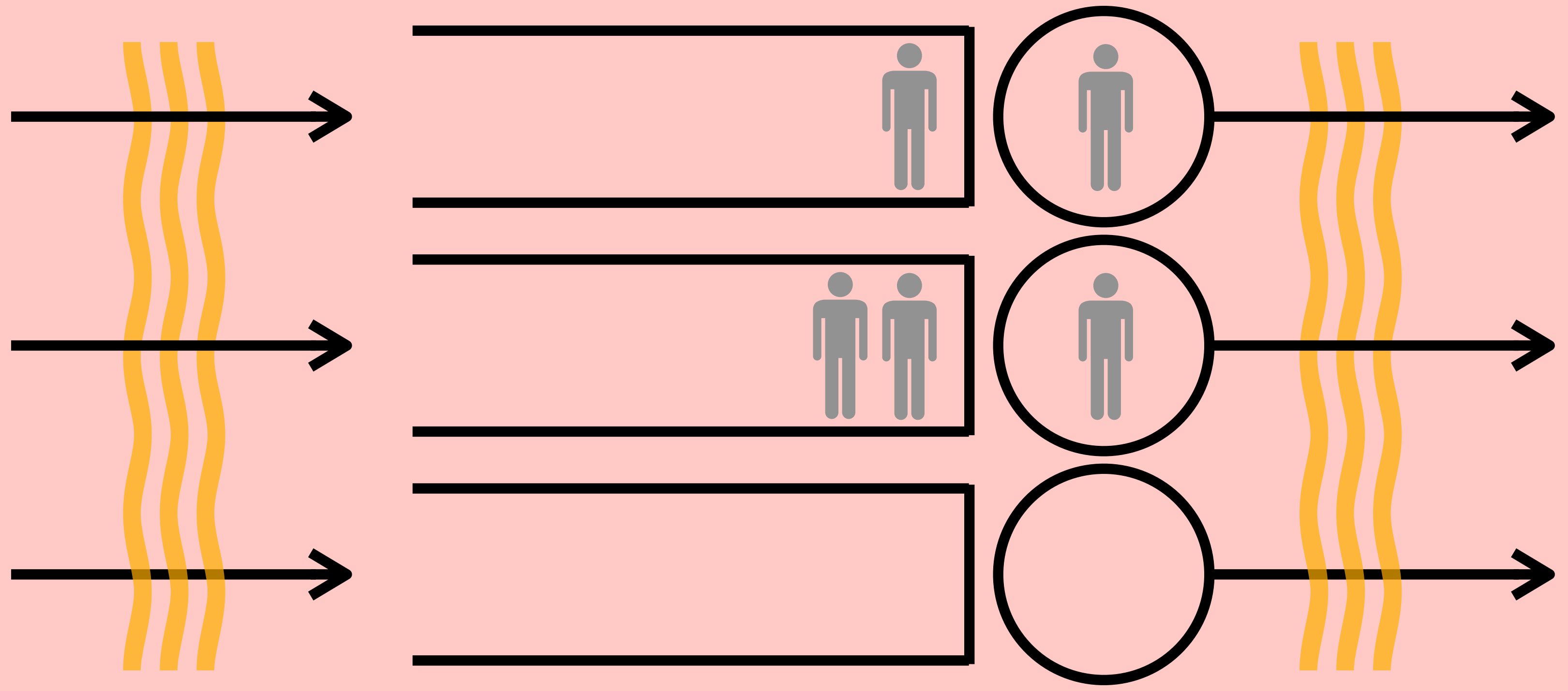
Tail prob: $P(\sum q_j > x)$?

Single-Server Queue



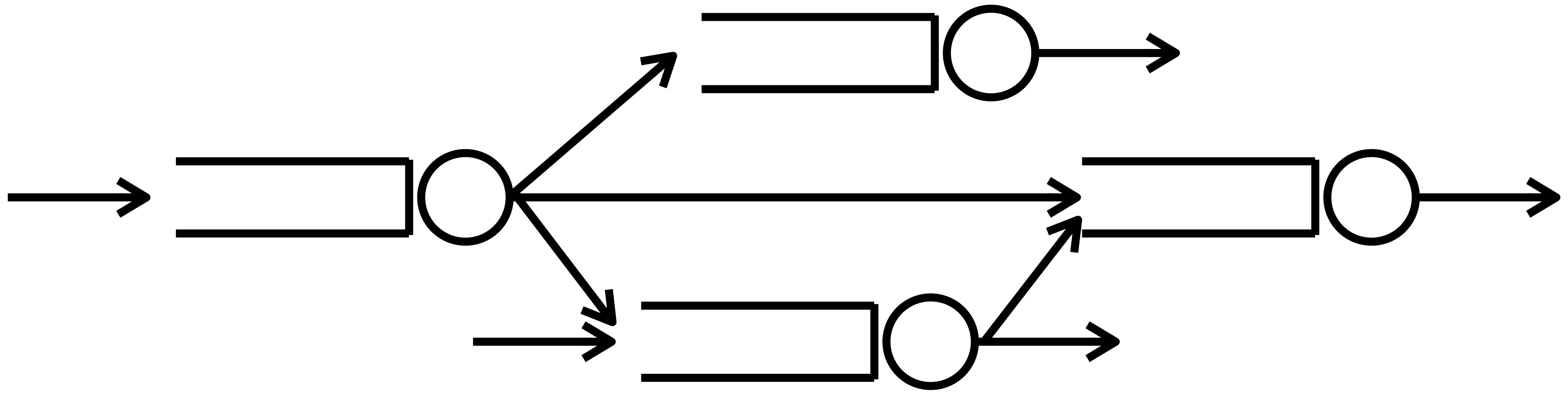
✓ Well understood
Distribution of q is known

Multiple queues with correlated arrivals and services



This talk

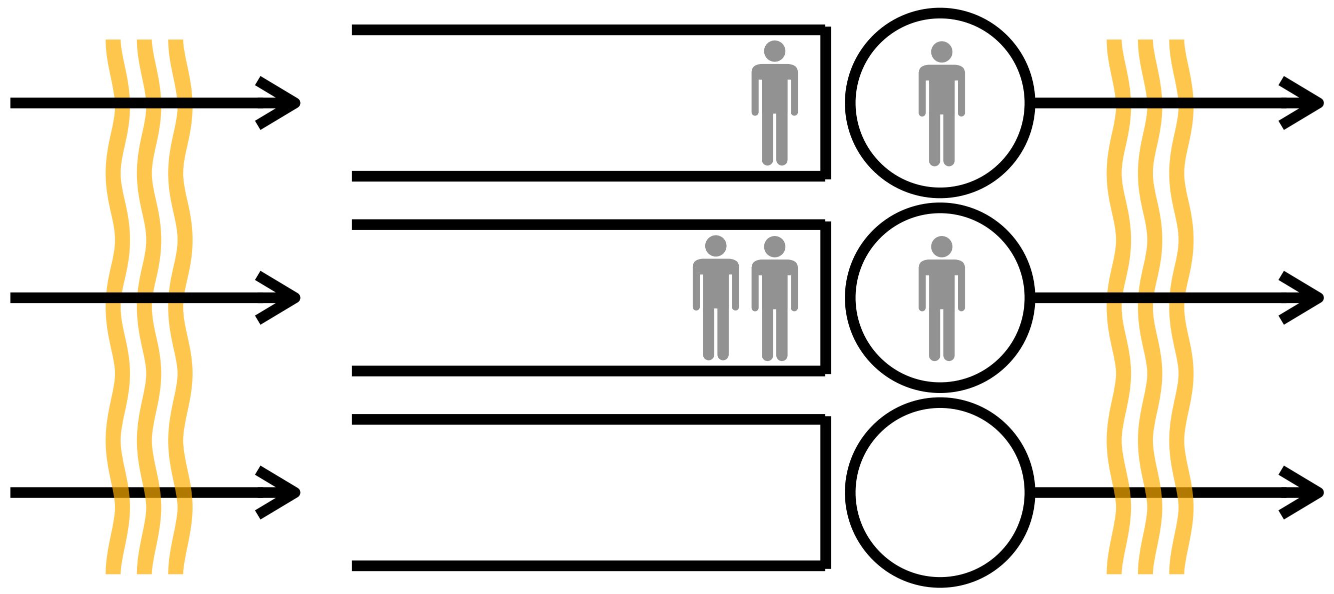
Networks of queues



✓ Jackson networks:
Distribution is product form

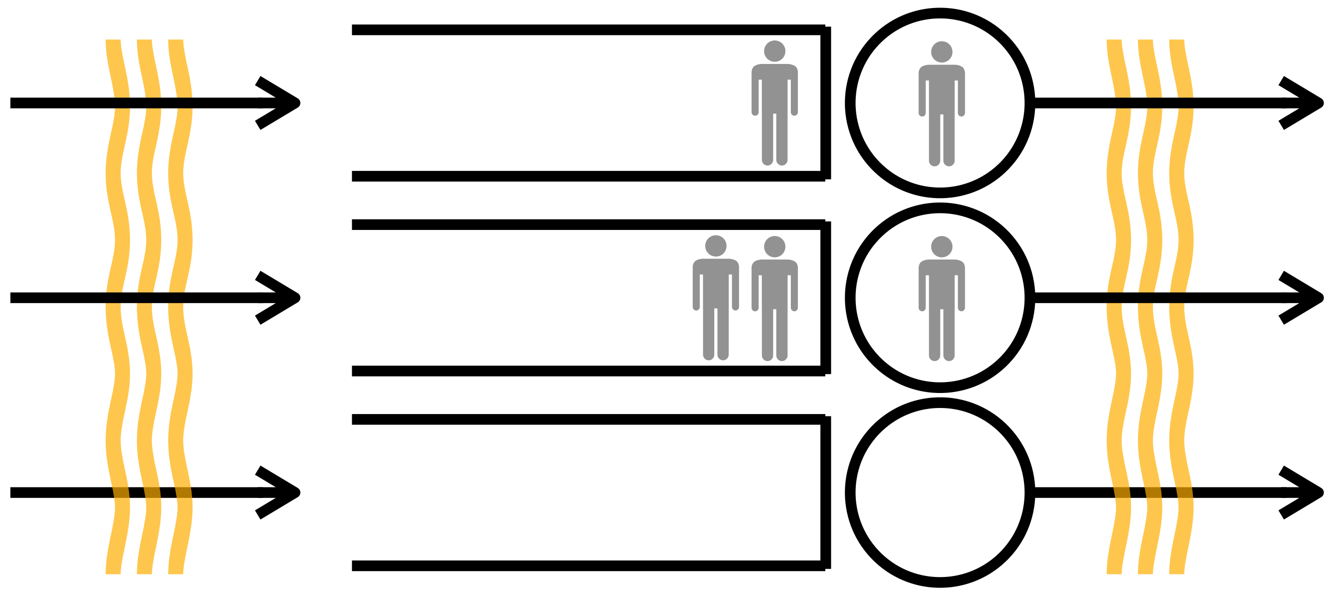
In the literature…

Asymptotic Regimes



In the literature…

Asymptotic Regimes



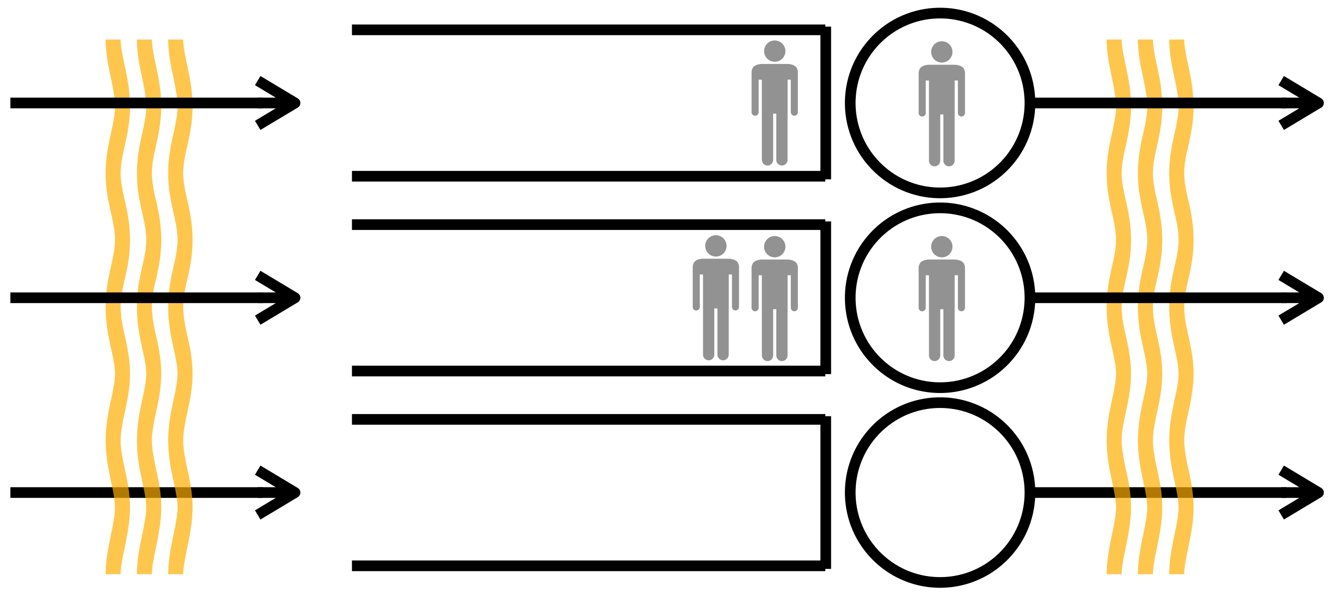
Large Deviations:
(a.k.a. rare events)

$$\lim_{x \to \infty} P(wait > x)$$

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

In the literature…

Asymptotic Regimes



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

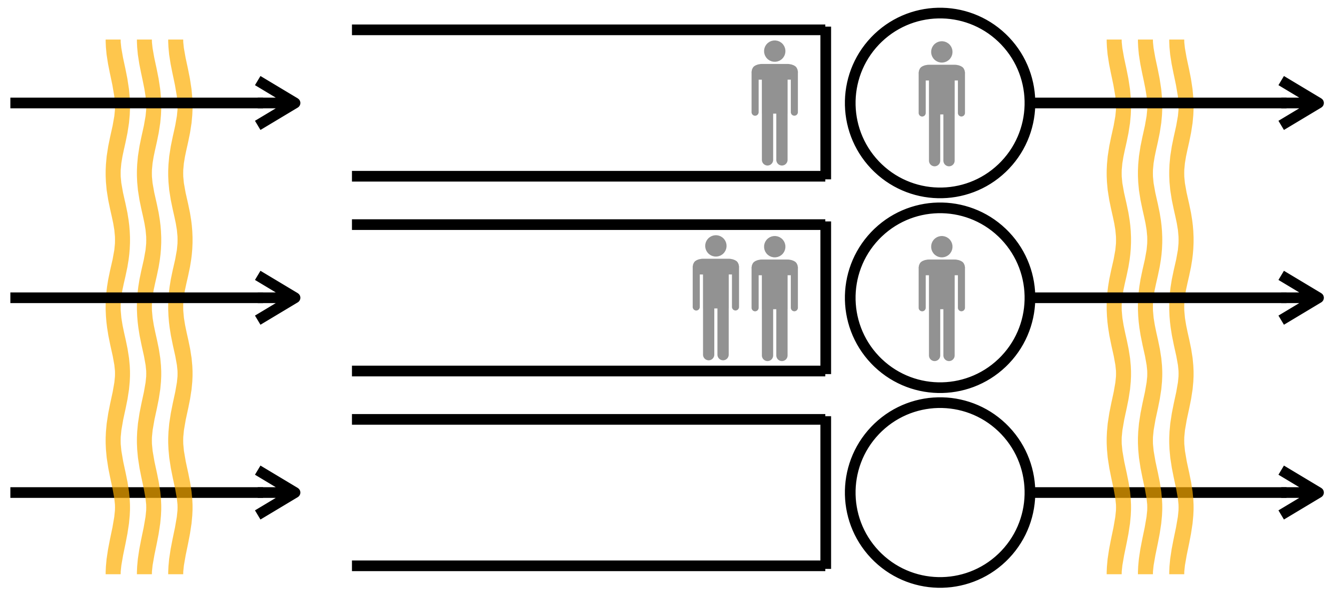
Ganesh (2004). Big Queues.
Glynn and Whitt (1994)
Puhalskii (1995)
Duffield and O'Connell (1995)
Dupuis and Ellis (1995)
…

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

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

In the literature…

Asymptotic Regimes



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

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 $\to \infty$

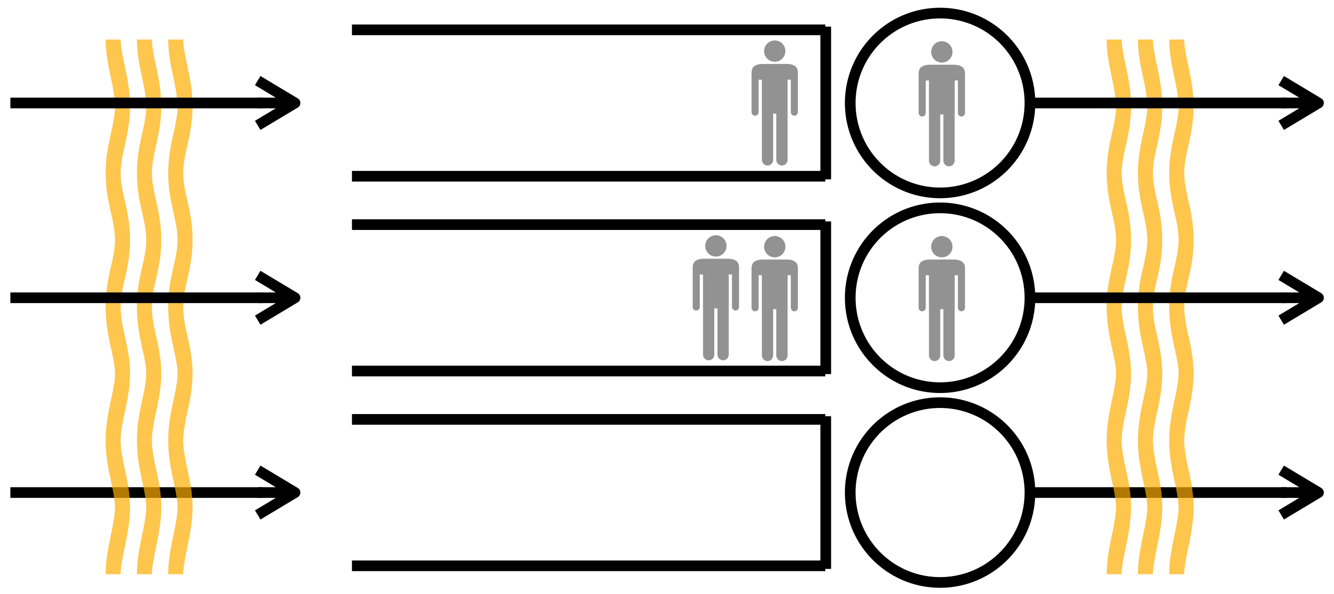
Ying (2017)
Mukkerjee et.al. (2018)
Mitzenmacher (1996, 2001)
Stolyar (2017)
…

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)
…

In the literature…

Asymptotic Regimes



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

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 $\to \infty$

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

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

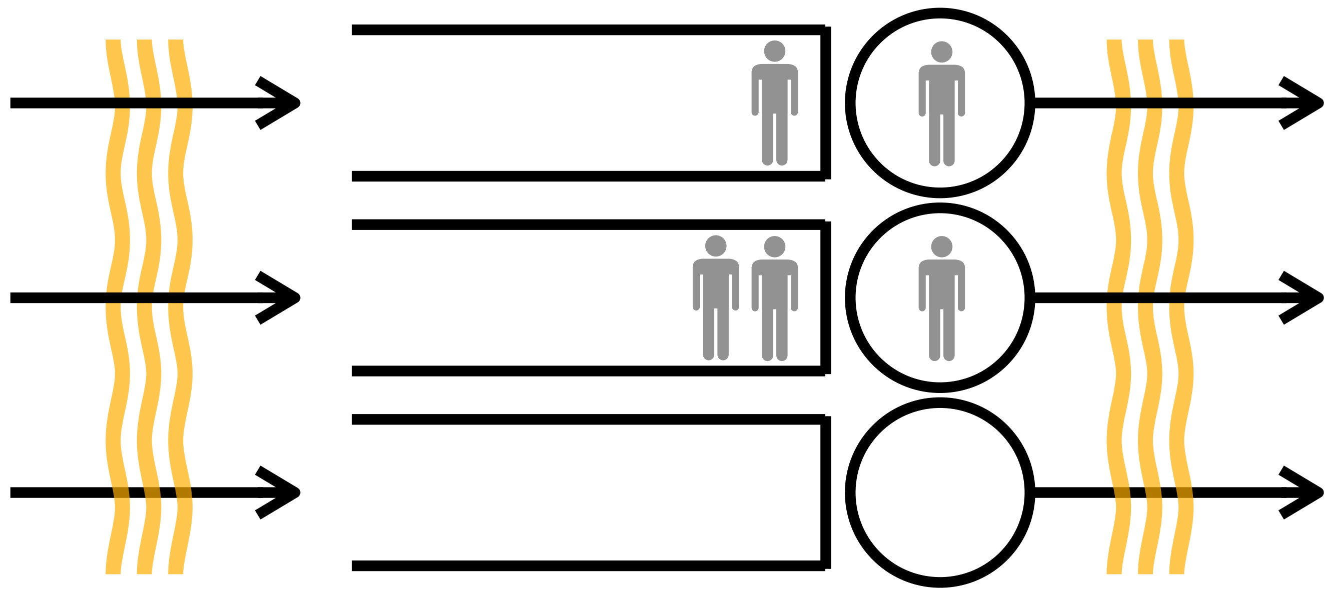
Banerjee et.al. (2019, 2020)
Liu and Ying (2019)
Braverman (2020)
Varma et.al. (2022)
HL, Maguluri (2022)
Jhunjhunwala, **HL**, Maguluri (2023)
…

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)
…

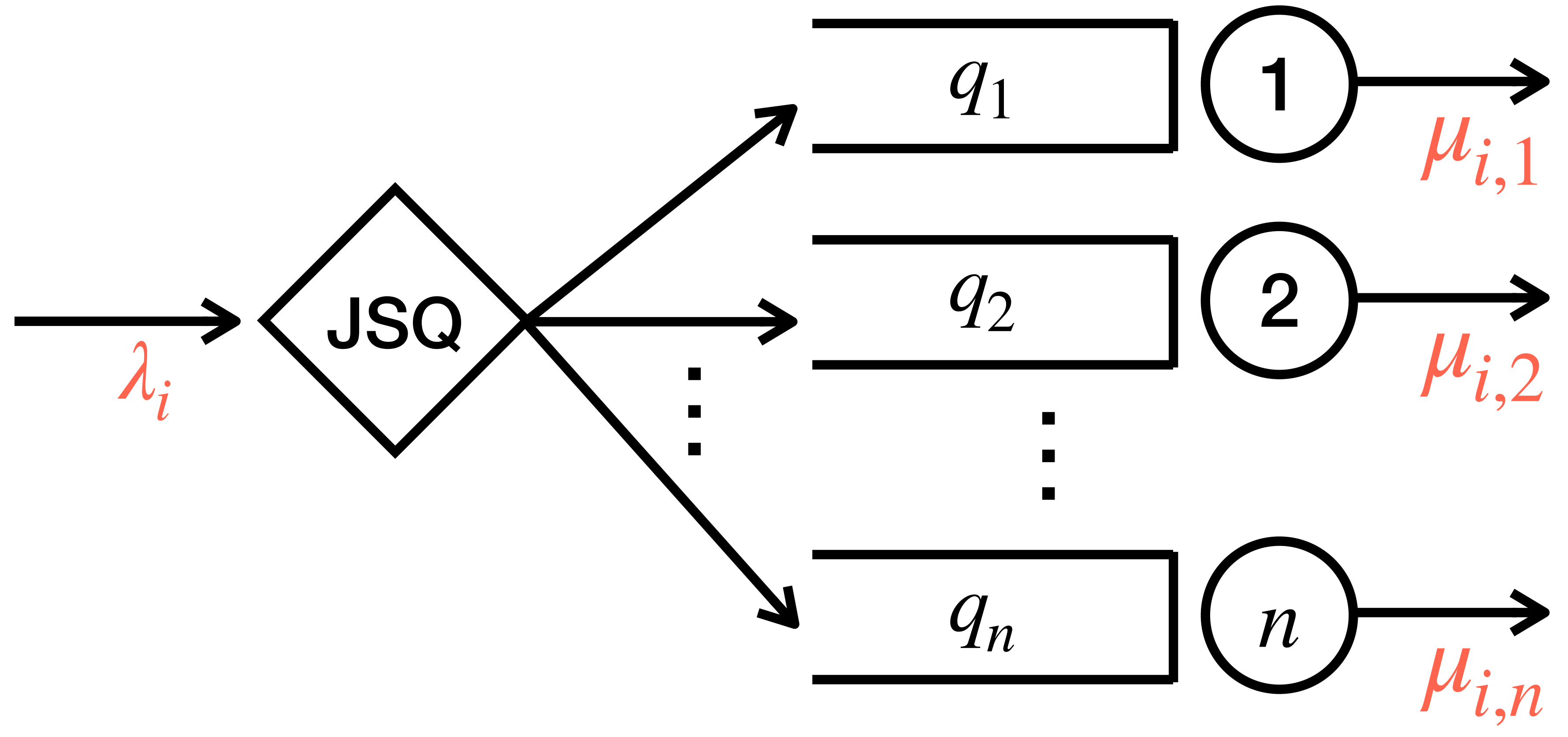
In the literature…

Asymptotic Regimes



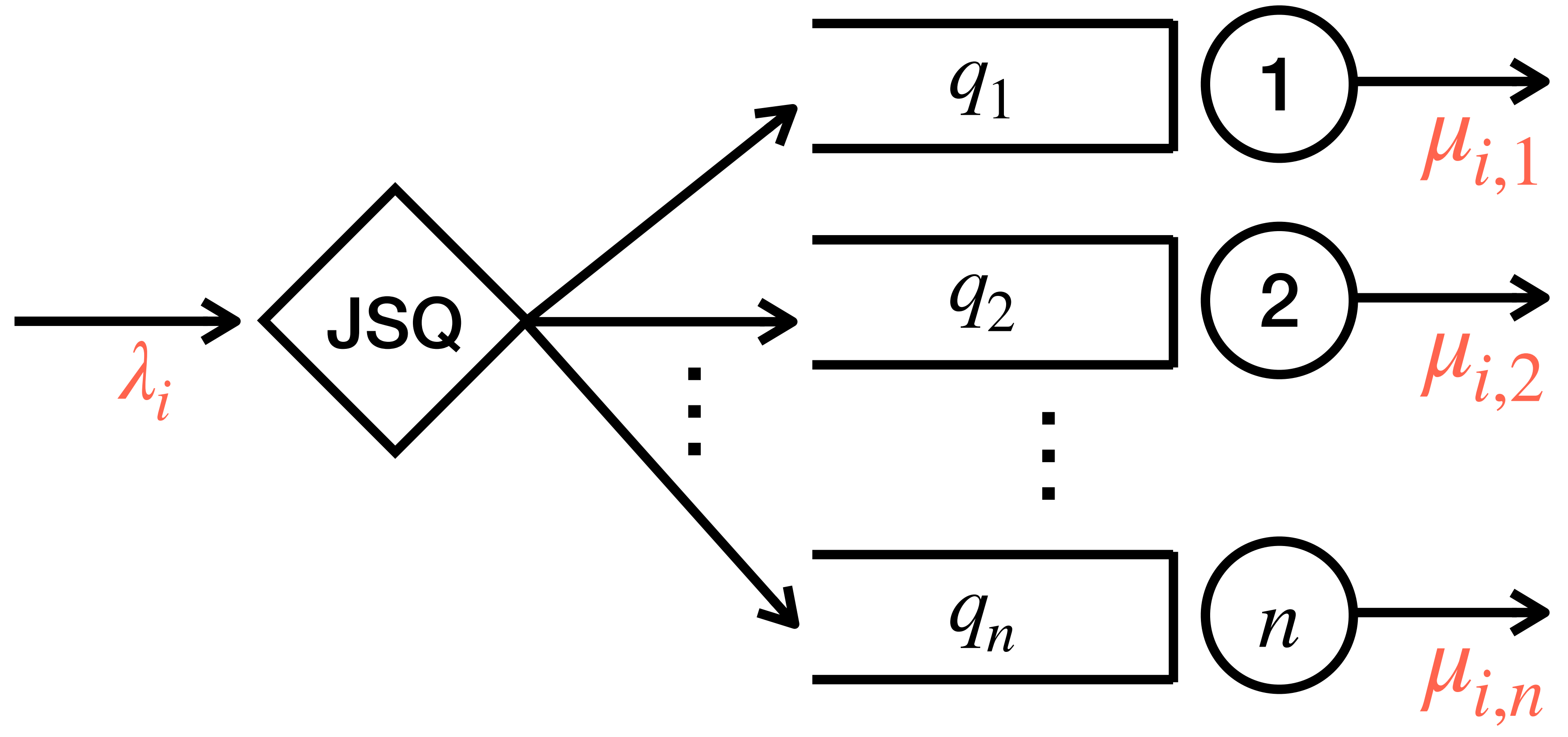
Large Deviations: (a.k.a. rare events) $\lim_{x\to\infty} P(wait > x)$	**Mean Field:** (a.k.a. many servers) # servers $\to \infty$	**Many-Server Heavy-Traffic:** # servers $\to \infty$ & Arrival rate $\approx$ Service rate	**Heavy Traffic:** Load system to max capacity $\iff$ Arrival rate $\approx$ Service rate
Ganesh (2004). Big Queues. Glynn and Whitt (1994) Puhalskii (1995) Duffield and O'Connell (1995) Dupuis and Ellis (1995) …	Ying (2017) Mukkerjee et.al. (2018) Mitzenmacher (1996, 2001) Stolyar (2017) …	Banerjee et.al. (2019, 2020) Liu and Ying (2019) Braverman (2020) Varma et.al. (2022) **HL**, Maguluri (2022) Jhunjhunwala, **HL**, Maguluri (2023) …	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) …

JSQ Model



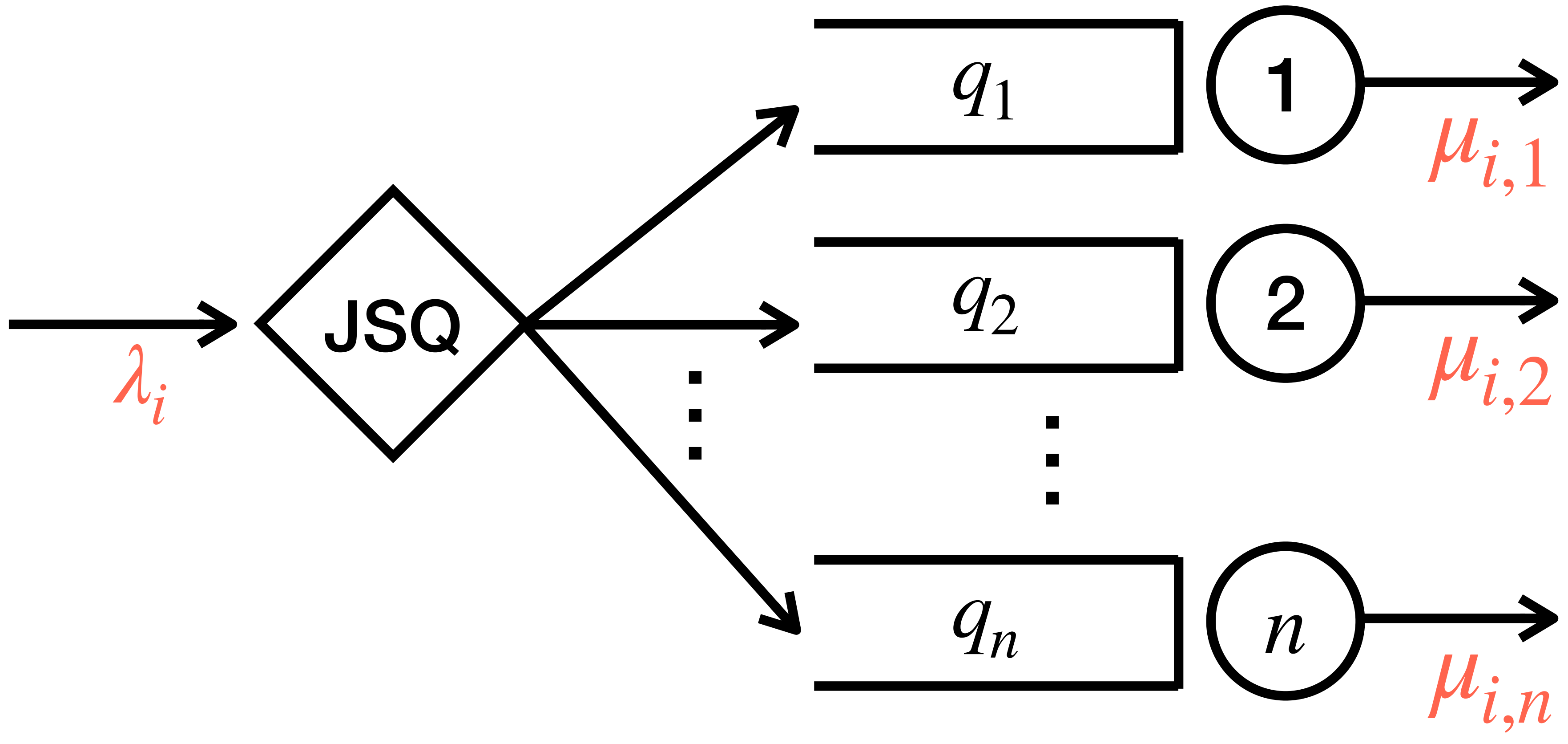
JSQ Model

- Load balancing under Join-the-Shortest-Queue (JSQ)
- Exponential inter-arrival and service times
- Heterogeneous servers

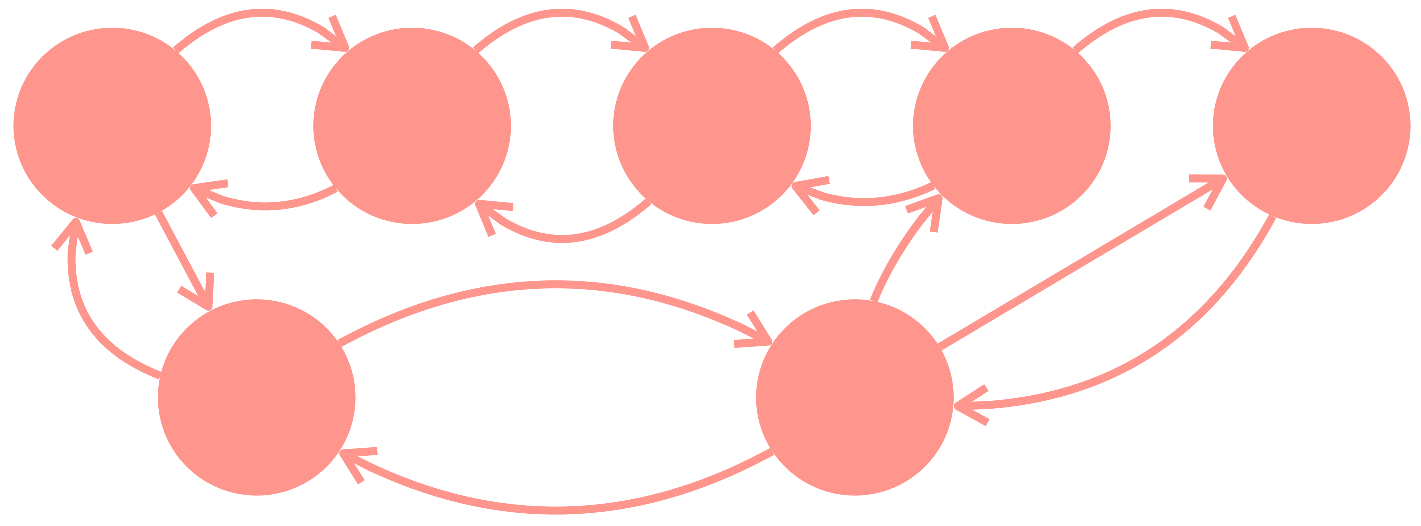


JSQ Model

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



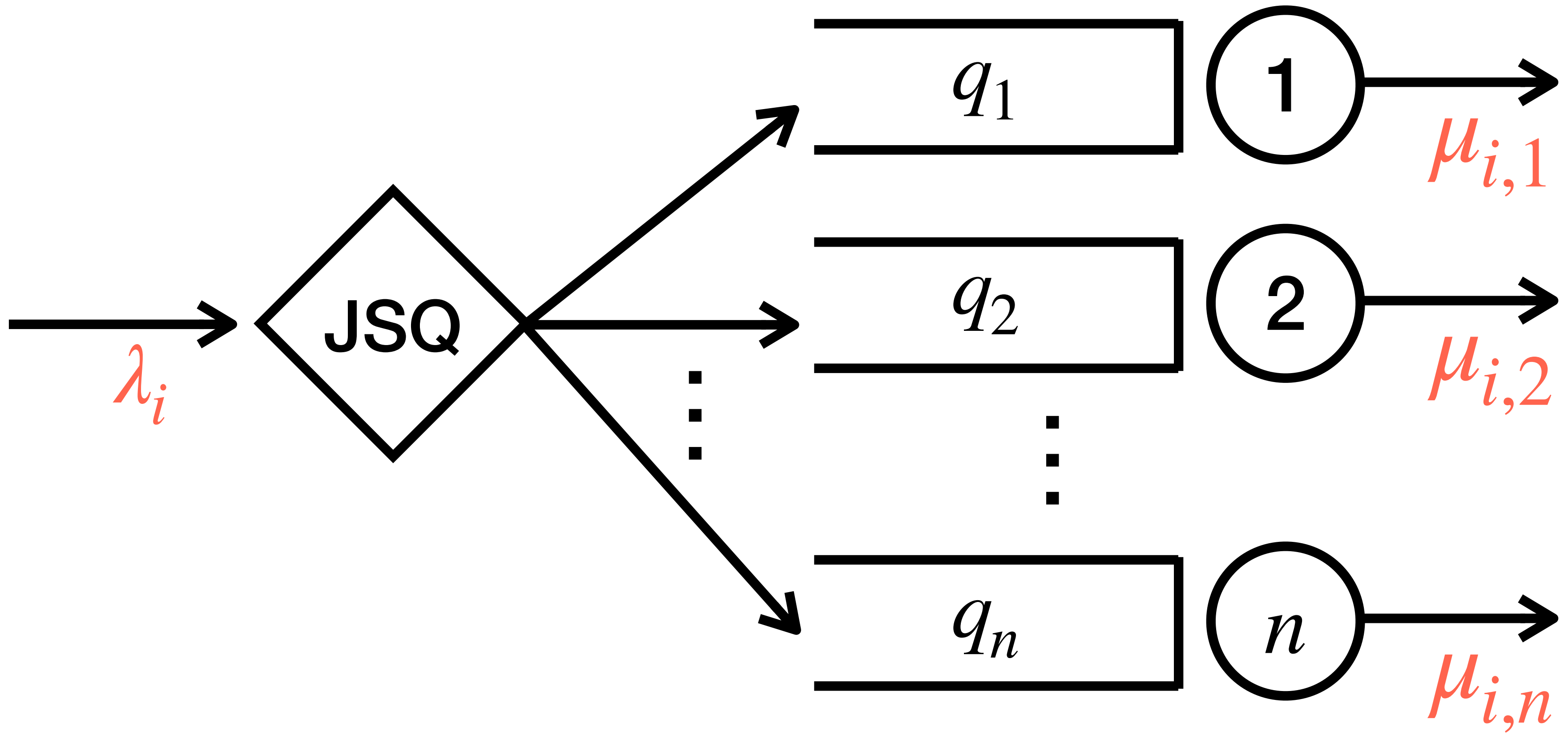
$Z(t) = i \sim$ Markov chain



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

JSQ Model

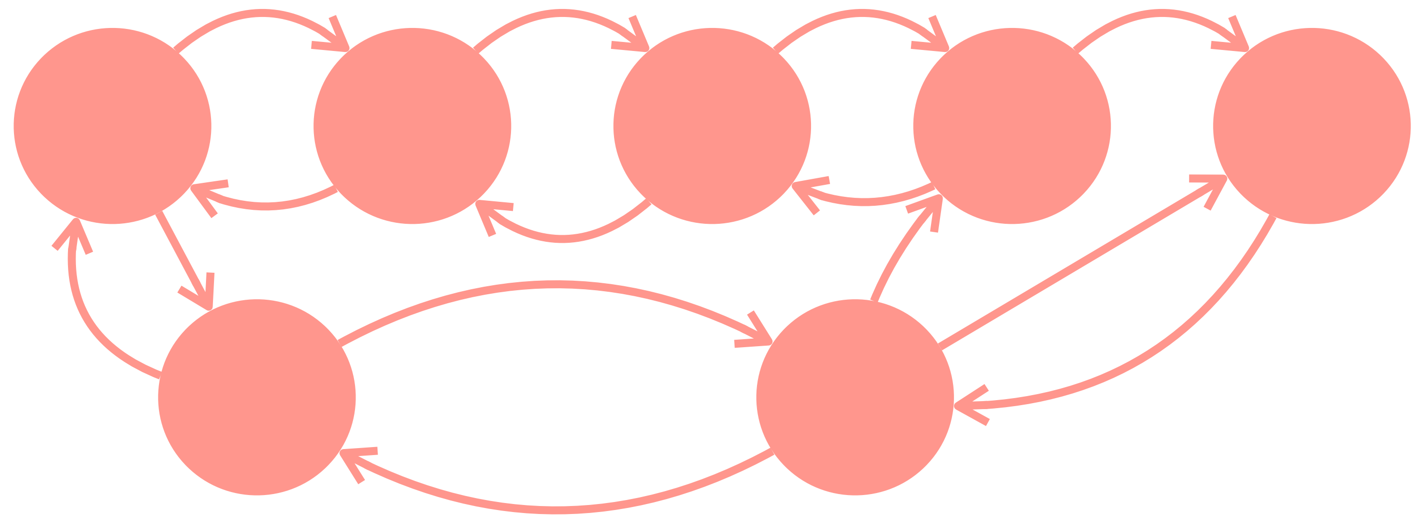
- Load balancing under Join-the-Shortest-Queue (JSQ)
- Exponential inter-arrival and service times
- Heterogeneous servers
- Arrival and service rate are Markov-Modulated



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

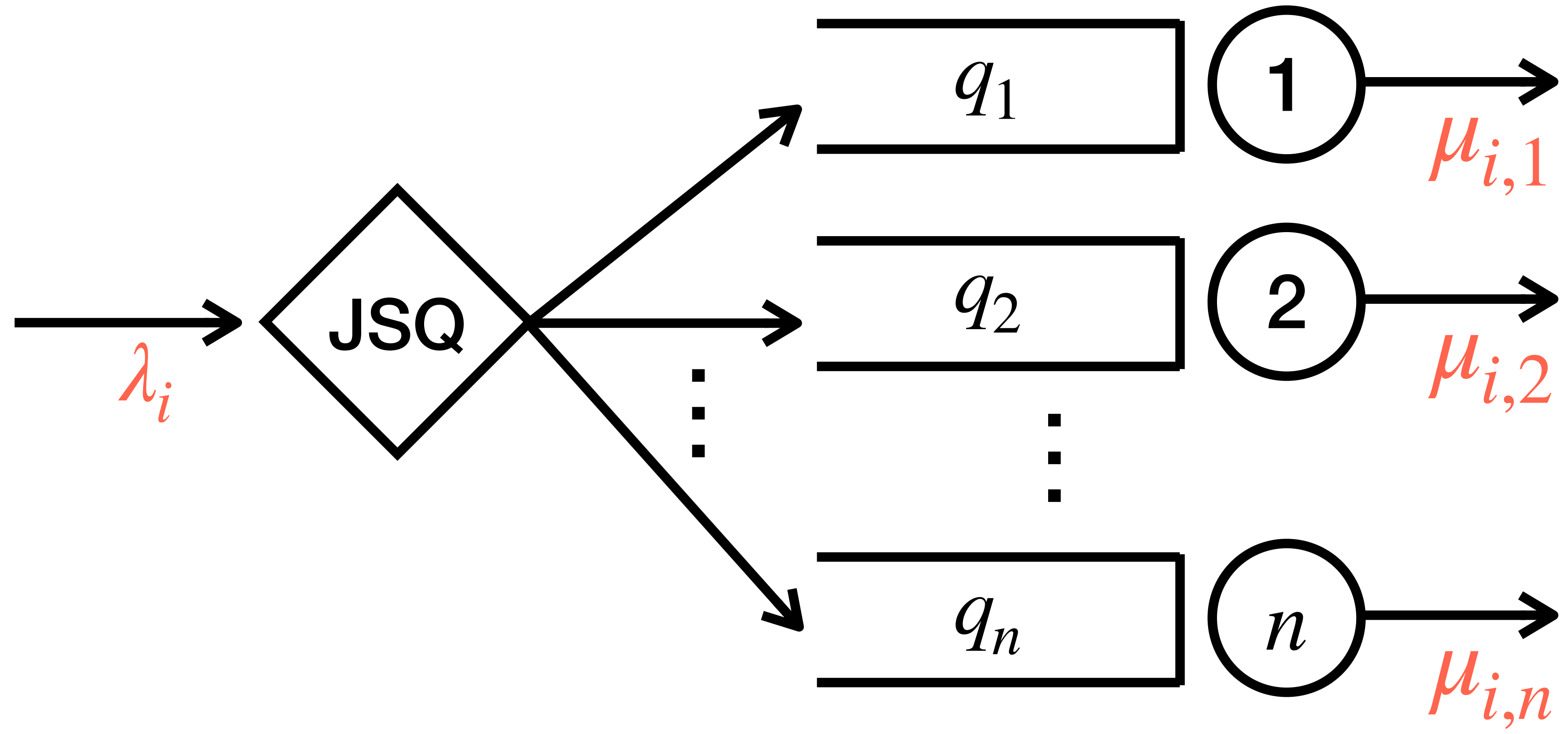
- Countable state space
- Stationary distribution exists
- $\lambda_{\max} < \infty$ and $\mu_{\max} < \infty$

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

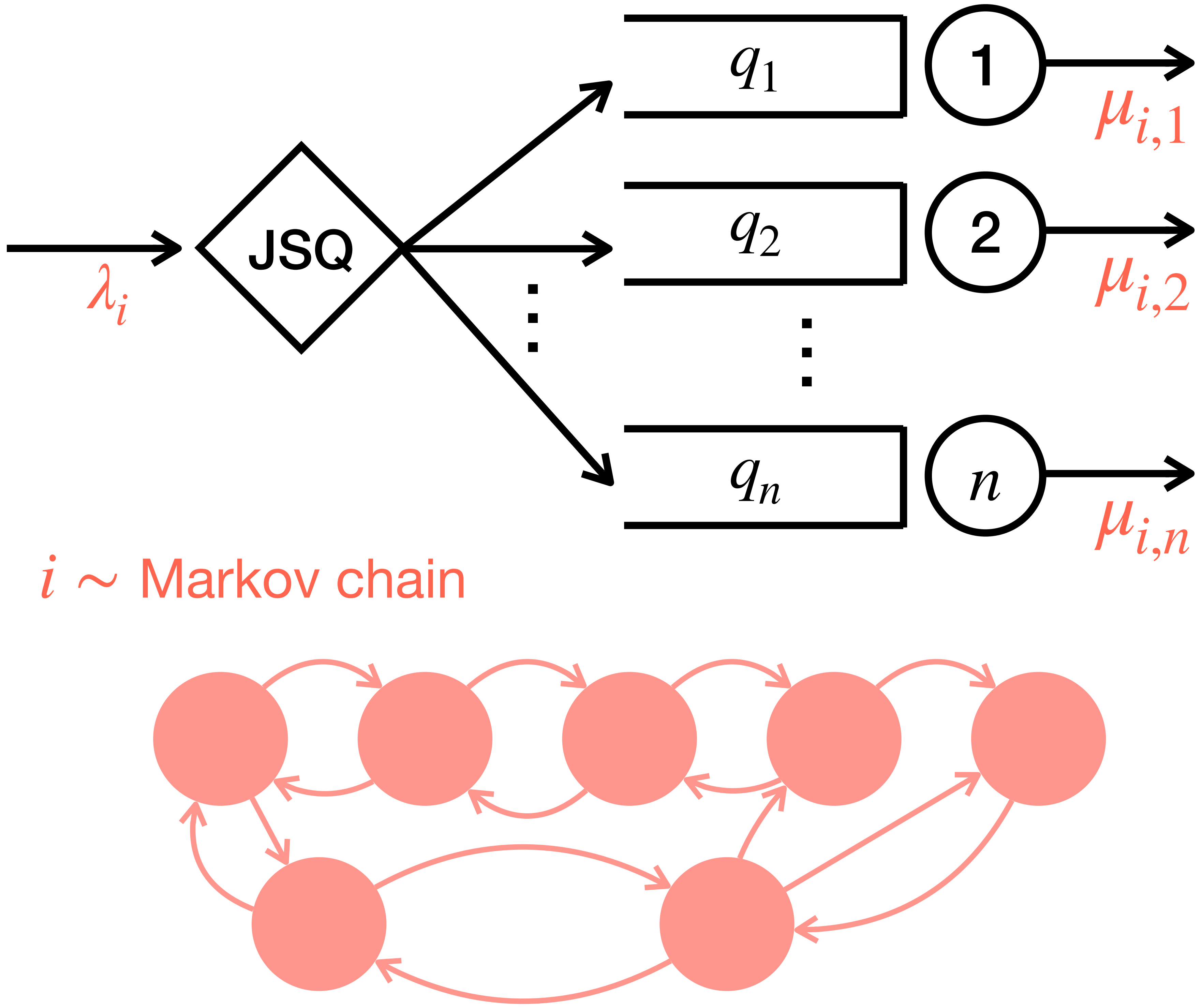


$$\rho_\pi = \frac{\mathbb{E}[\lambda_i \,|\, i \sim \pi]}{\sum_j \mathbb{E}[\mu_{ij} \,|\, i \sim \pi]}, \quad \epsilon = 1 - \rho_\pi$$

Asymptotic Distribution of Queue Lengths



Asymptotic Distribution of Queue Lengths



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

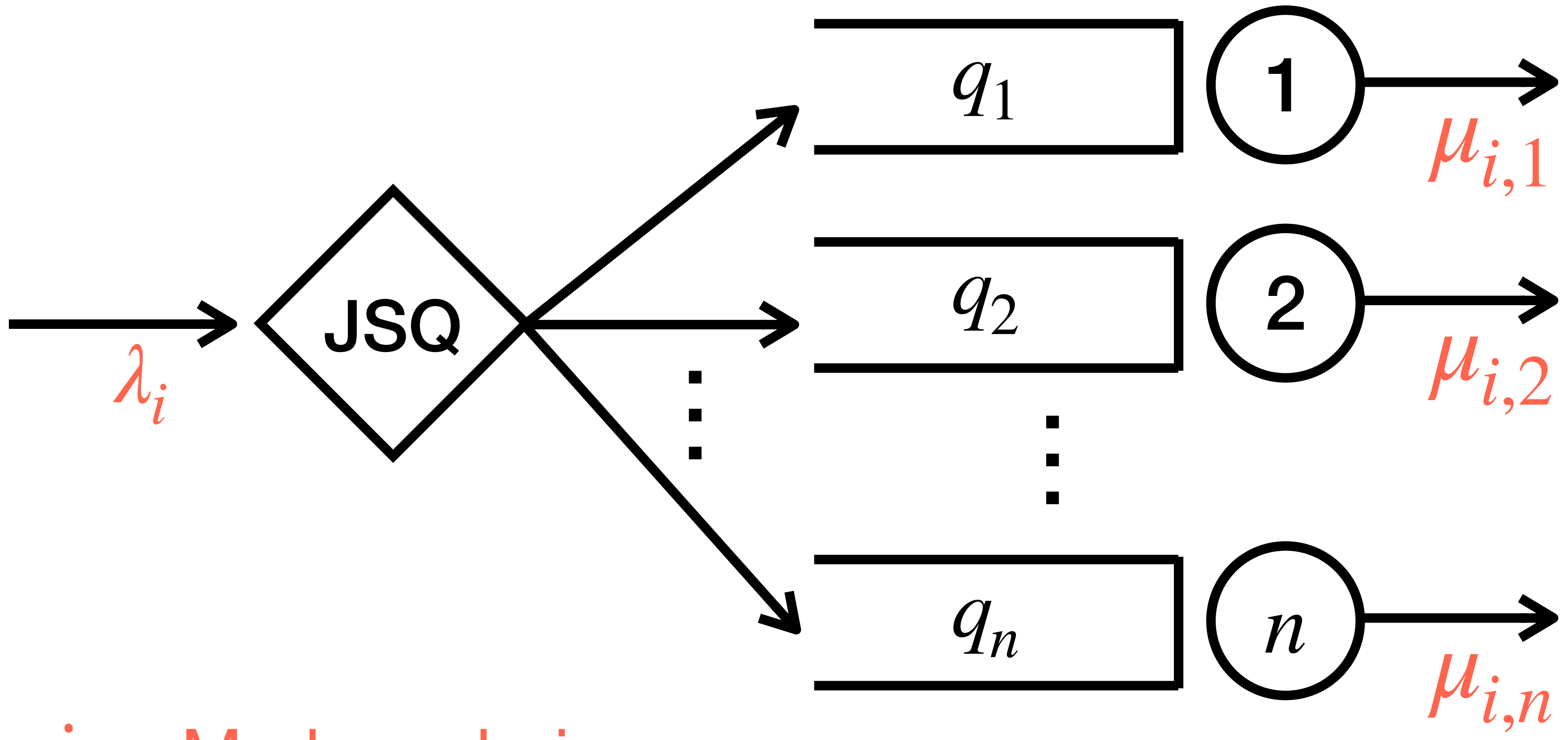
Asymptotic Distribution of Queue Lengths

Theorem [HL, Grosof '25]**:**

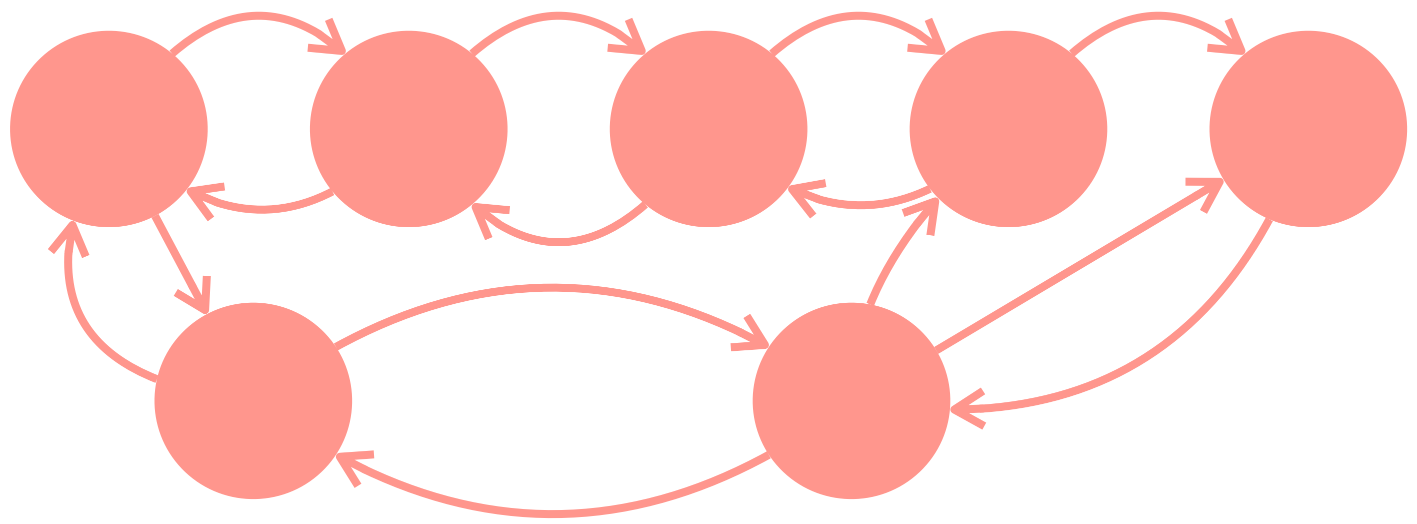
If $\lambda_i > 0$ is large enough for each i, then as $\epsilon \downarrow 0$

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

≈ Variance of arrival and service times in steady state



$i \sim$ Markov chain



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

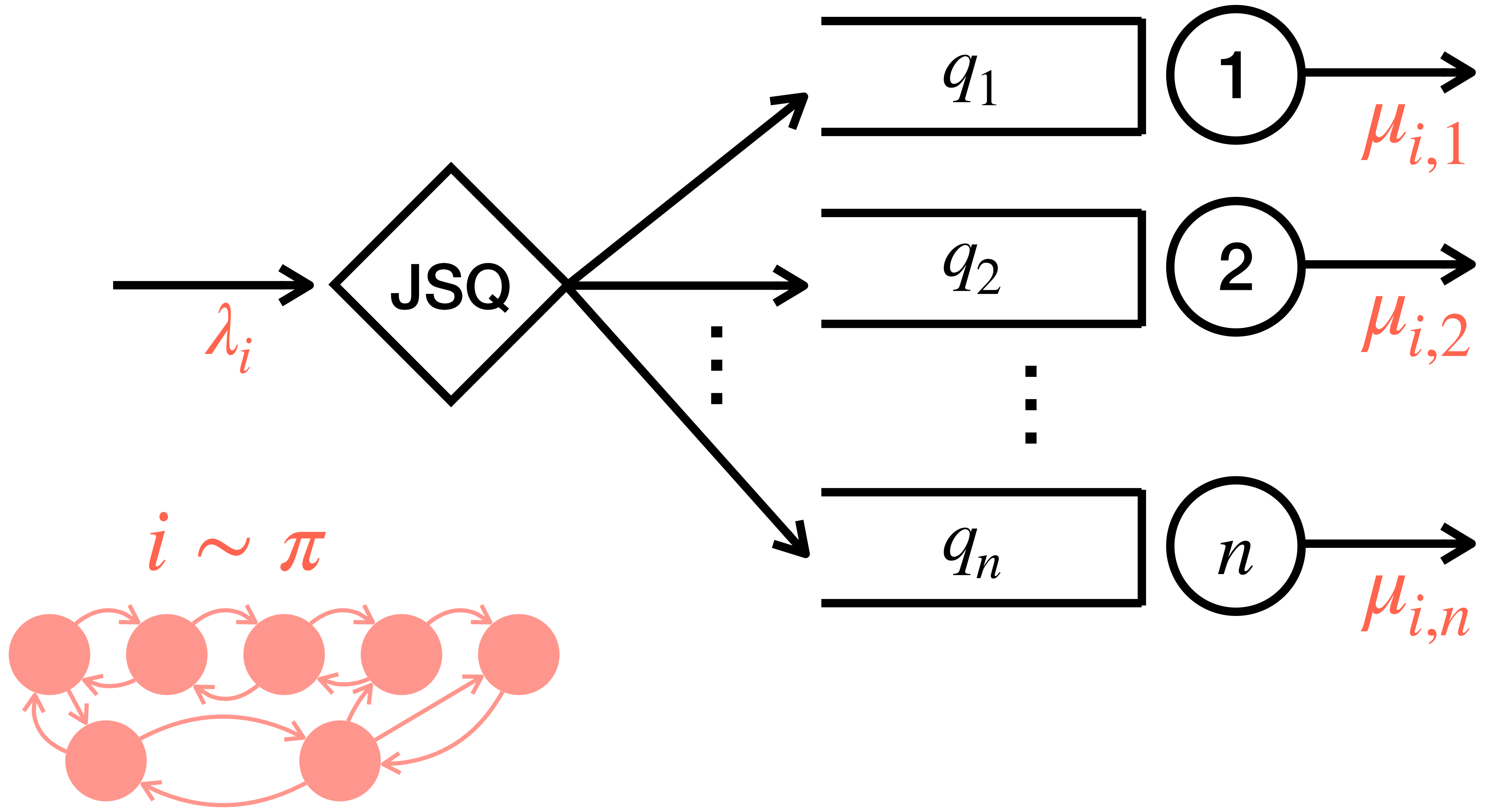
Proof Sketch

Theorem [HL, Grosof '25]:

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

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

≈ Variance of arrival and service times in steady state



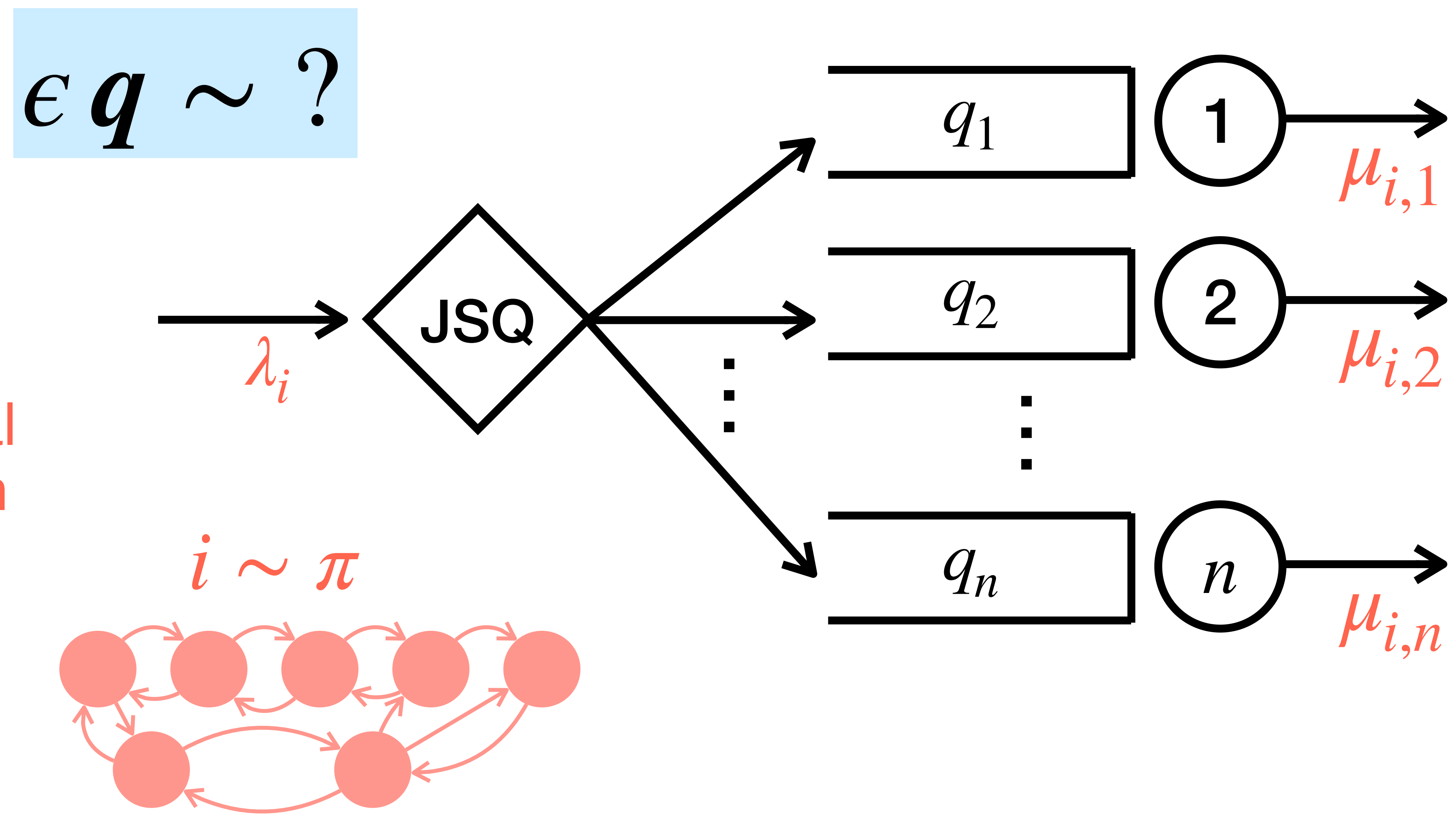
Proof Sketch

Theorem [HL, Grosof '25]:

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

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

$\approx$ Variance of arrival and service times in steady state



Proof Sketch

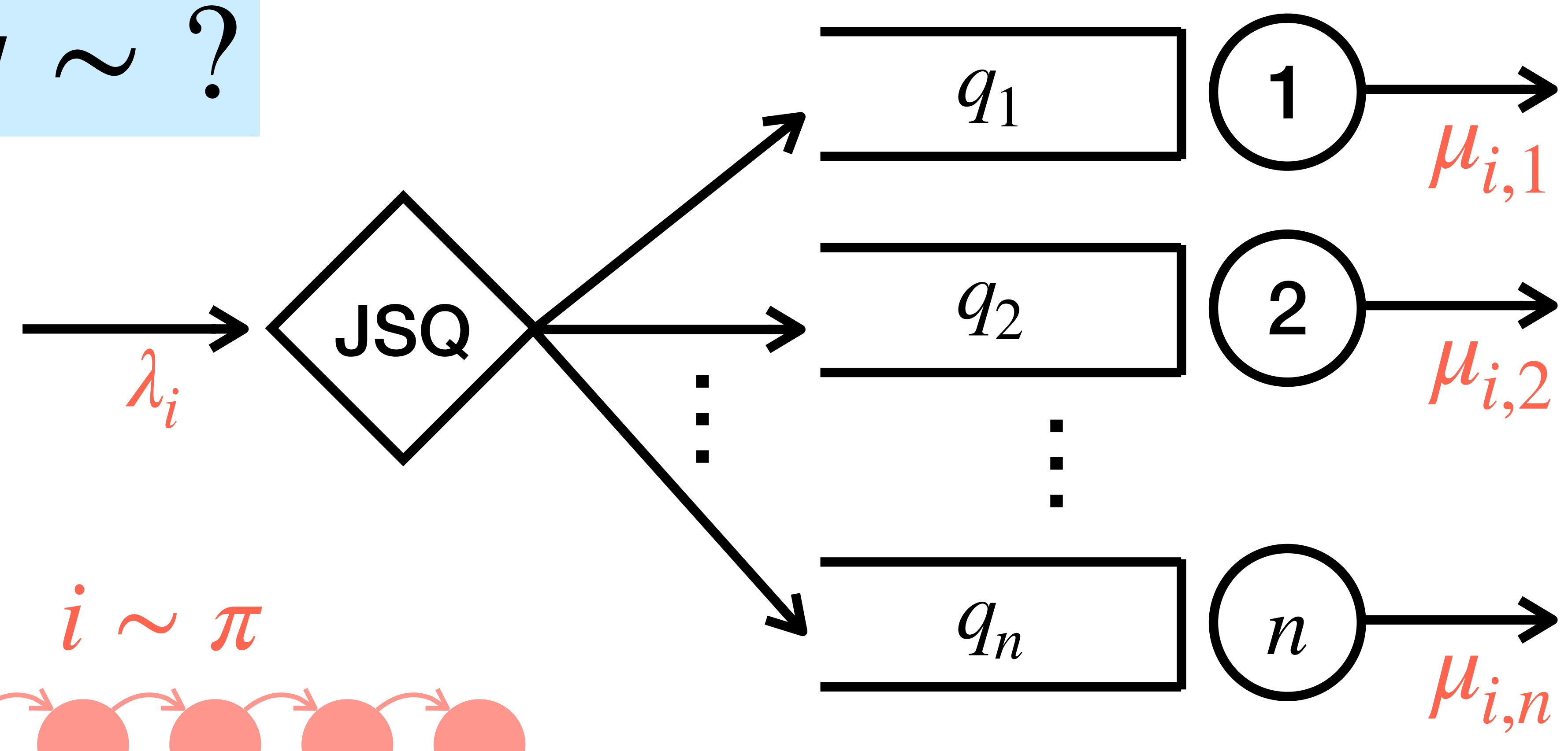
Theorem [HL, Grosof '25]:

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

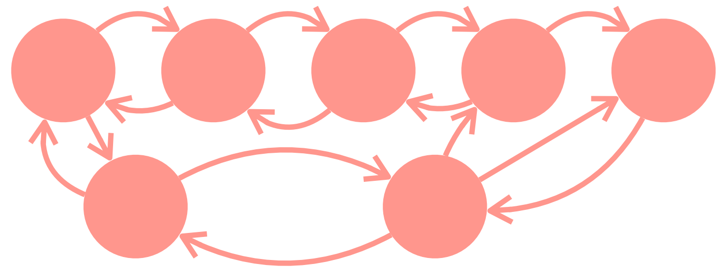
$$\epsilon \boldsymbol{q} \Longrightarrow Exp\left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma}\right)$$

$\approx$ Variance of arrival and service times in steady state

$\epsilon \boldsymbol{q} \sim ?$

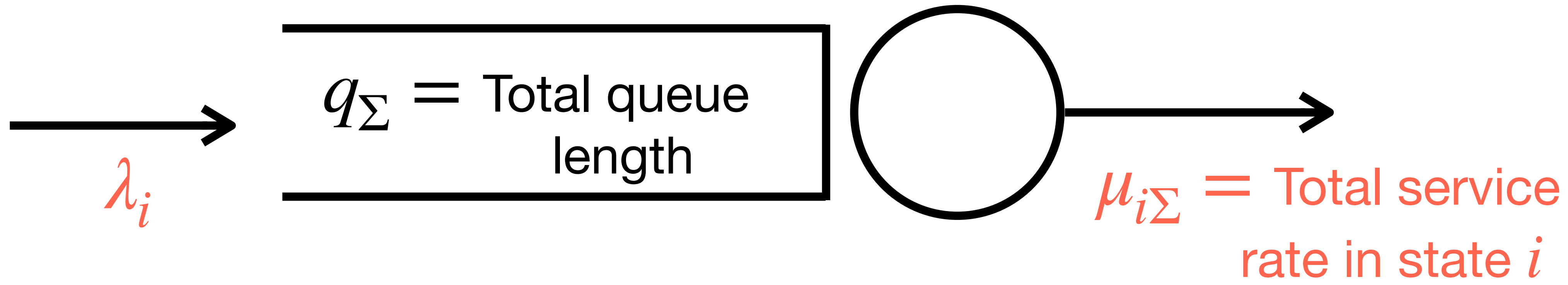


$i \sim \pi$



Step 1: State Space Collapse (SSC)

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



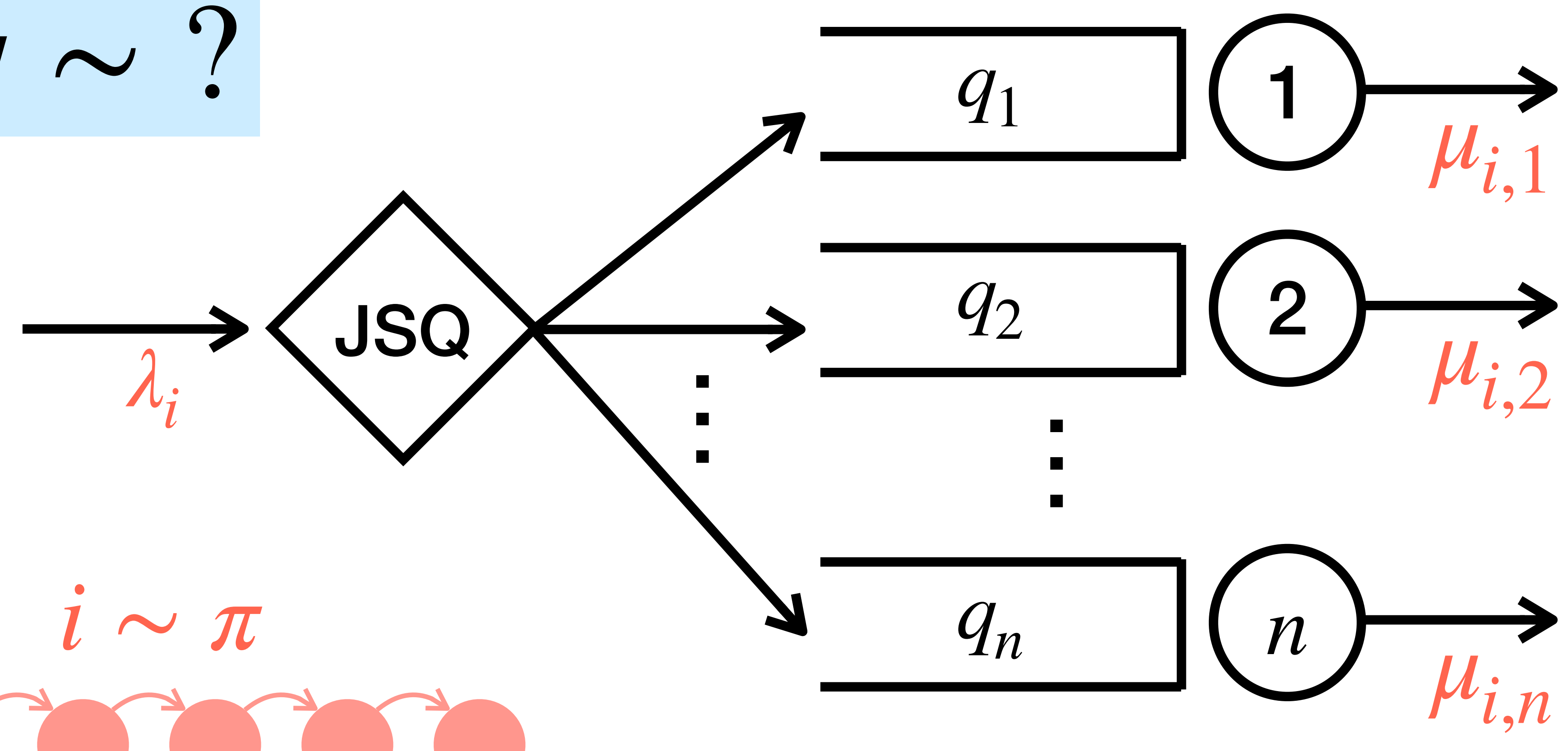
Proof Sketch

Theorem [HL, Grosof '25]:
If $\lambda_i > 0$ is large enough for each i, then as $\epsilon \downarrow 0$

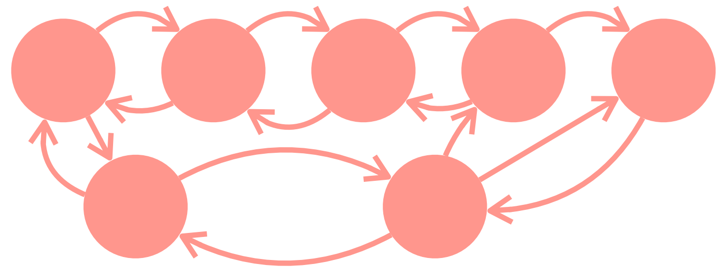
$$\epsilon\, \boldsymbol{q} \Longrightarrow Exp\left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma}\right)$$

$\approx$ Variance of arrival and service times in steady state

$\epsilon\, \boldsymbol{q} \sim ?$

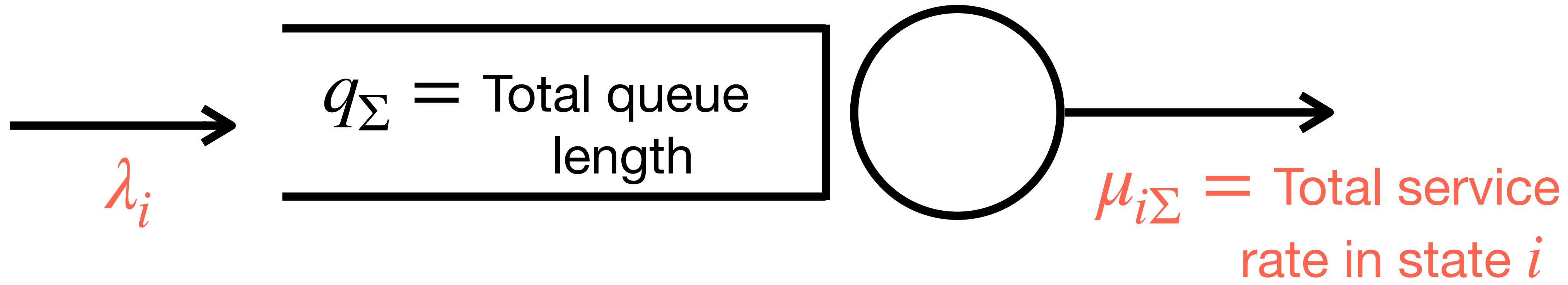


$i \sim \pi$



Step 1: State Space Collapse (SSC)

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



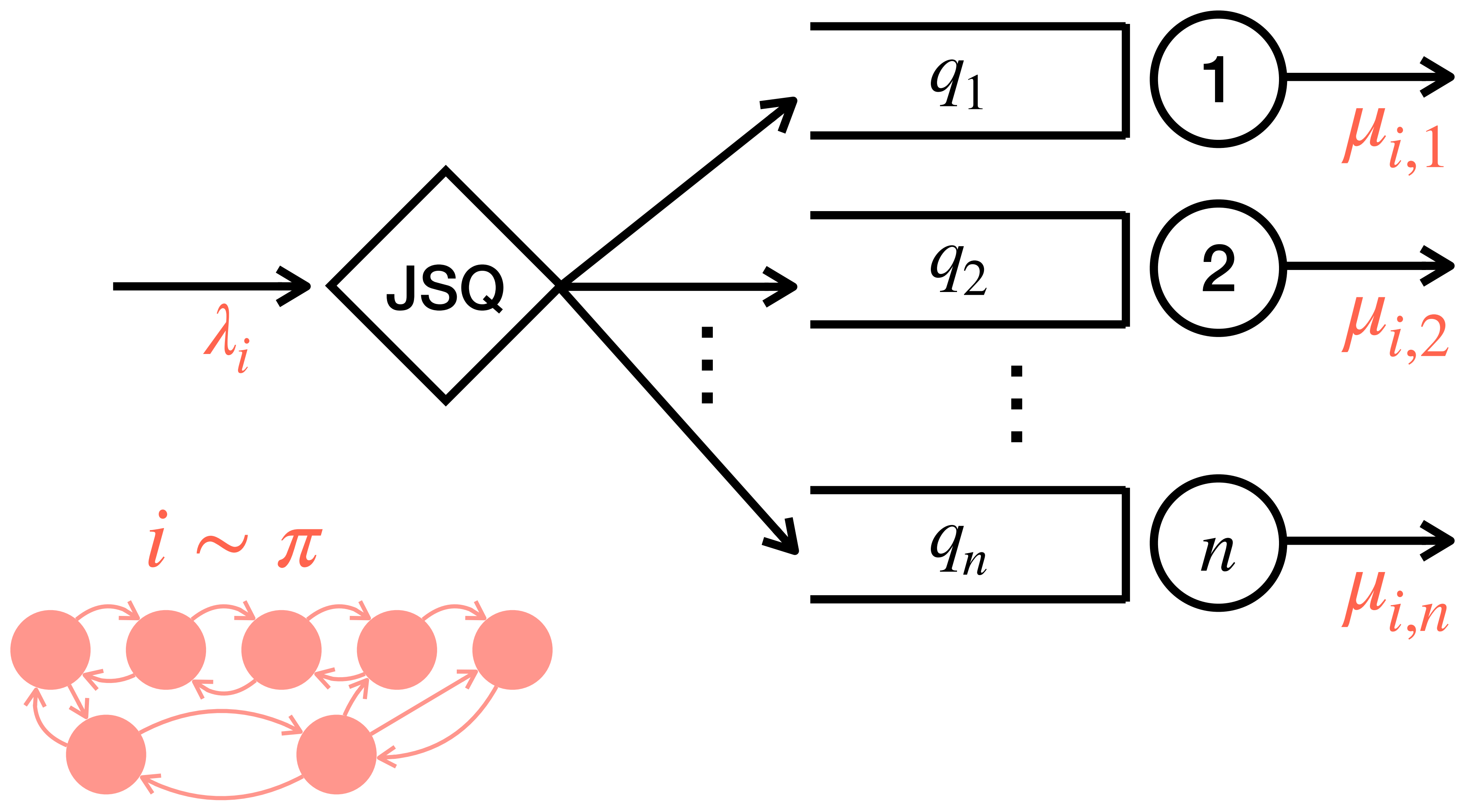
Step 2: Asymptotic distribution

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

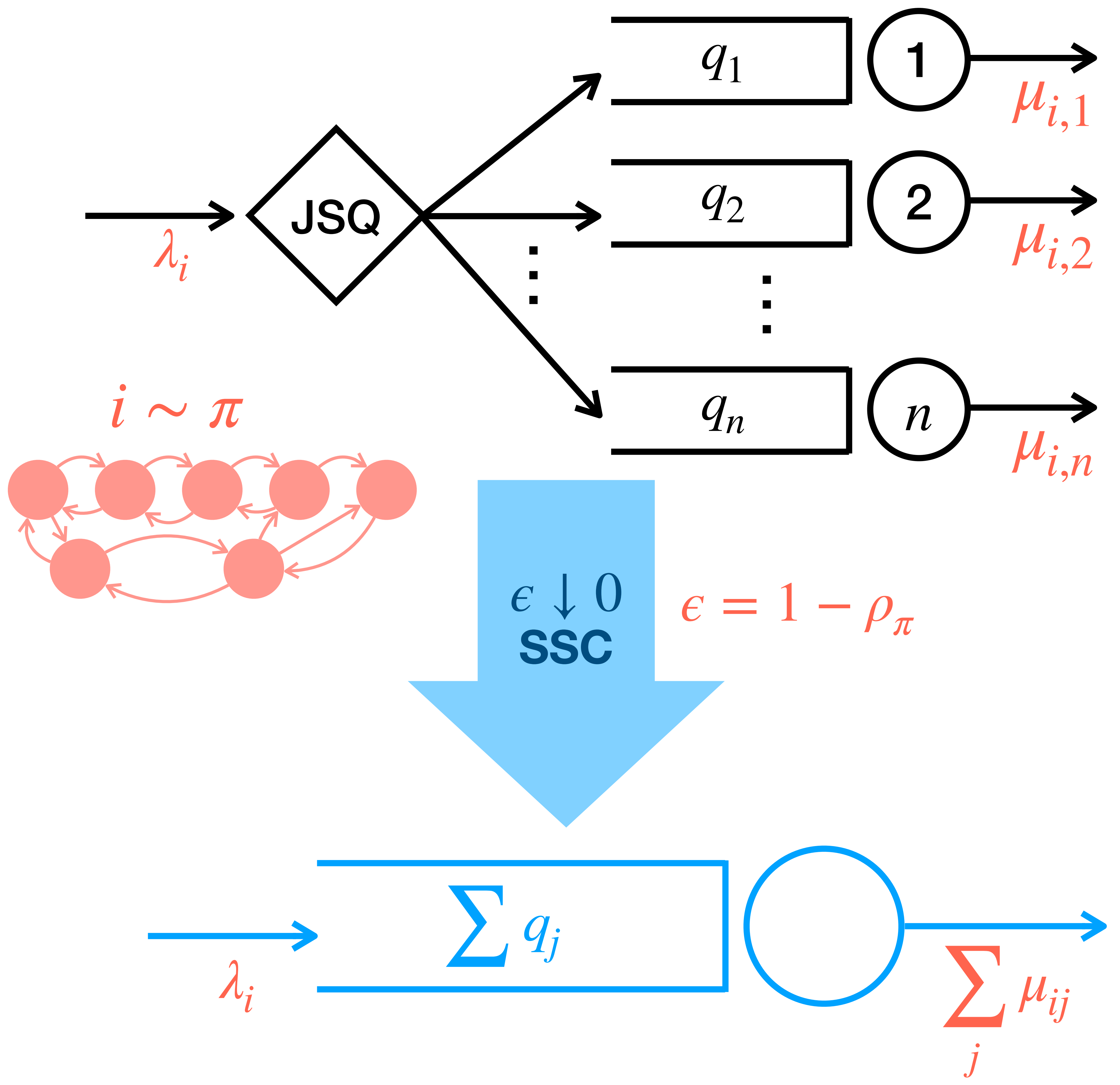
We use:

- Transform Method
- Poisson Equation

State Space Collapse (SSC)



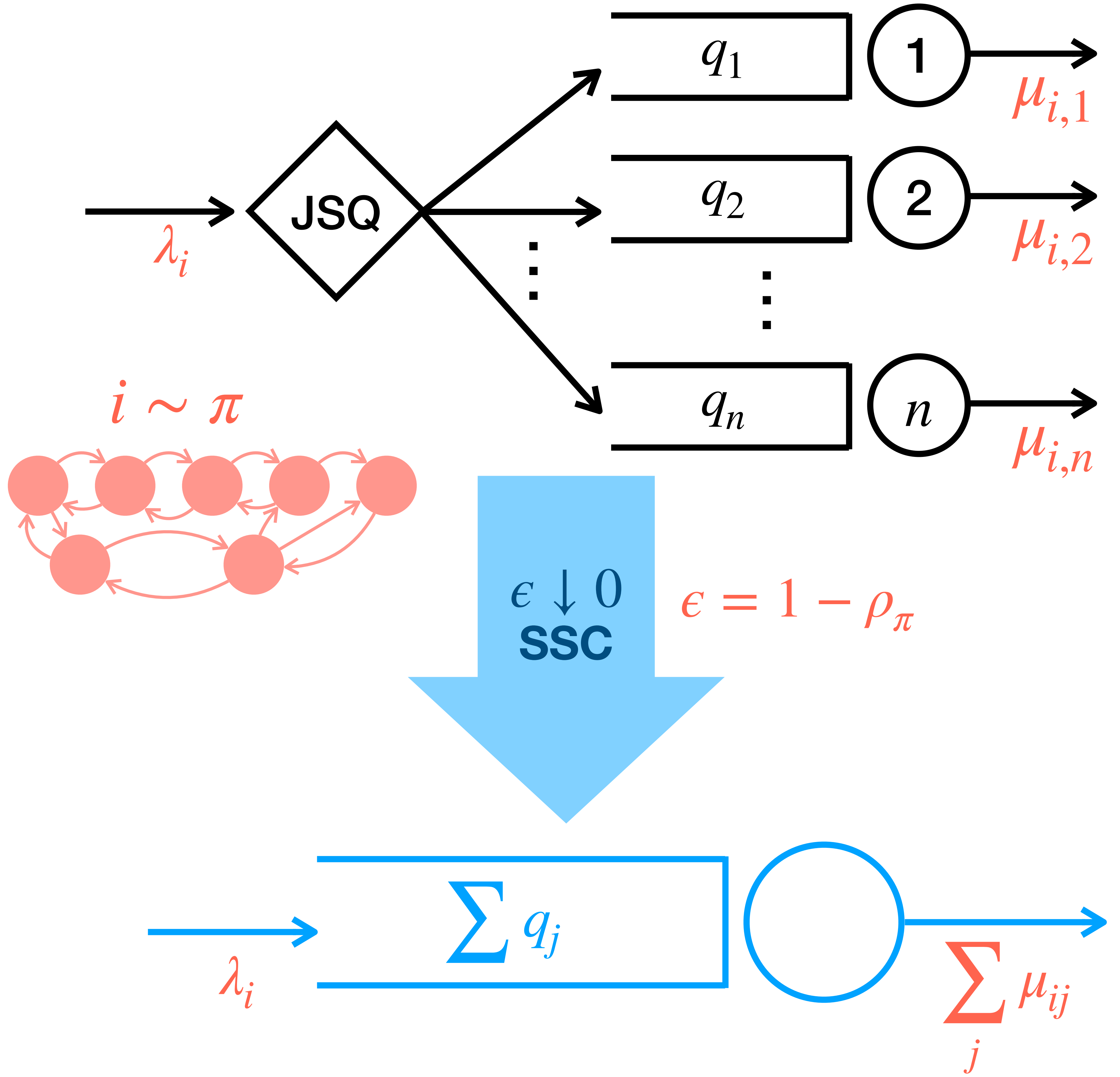
State Space Collapse (SSC)



State Space Collapse (SSC)

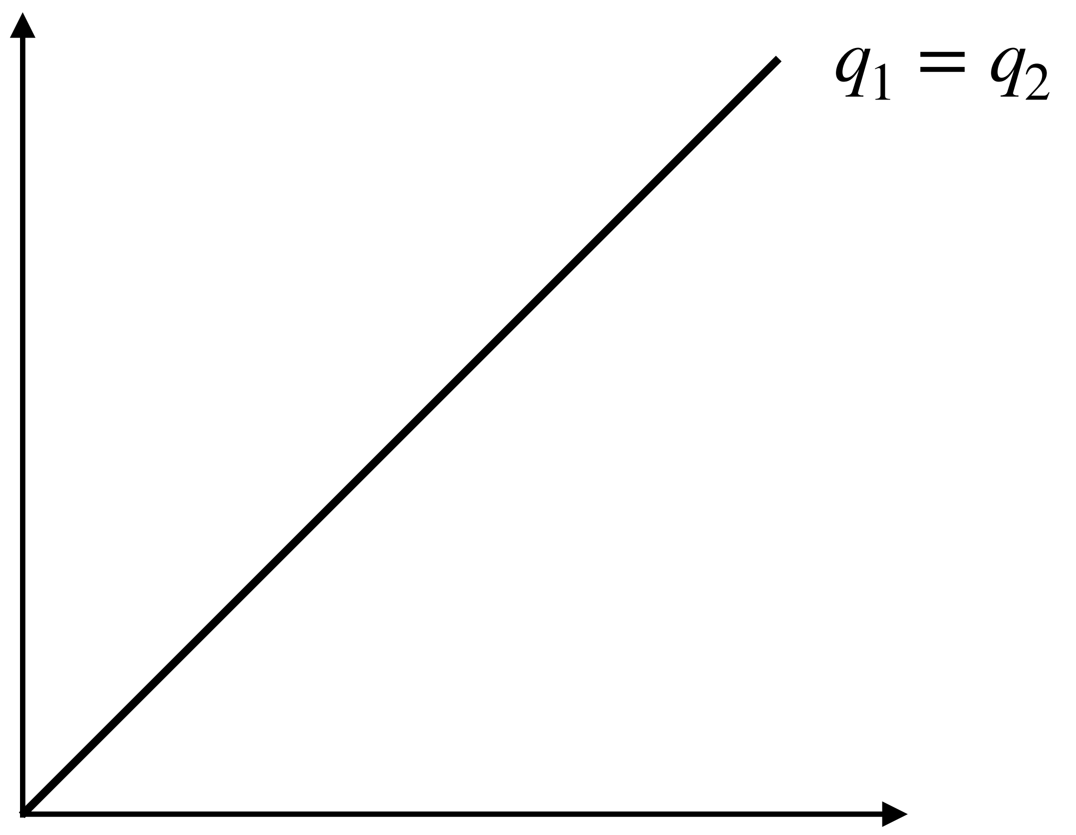
Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

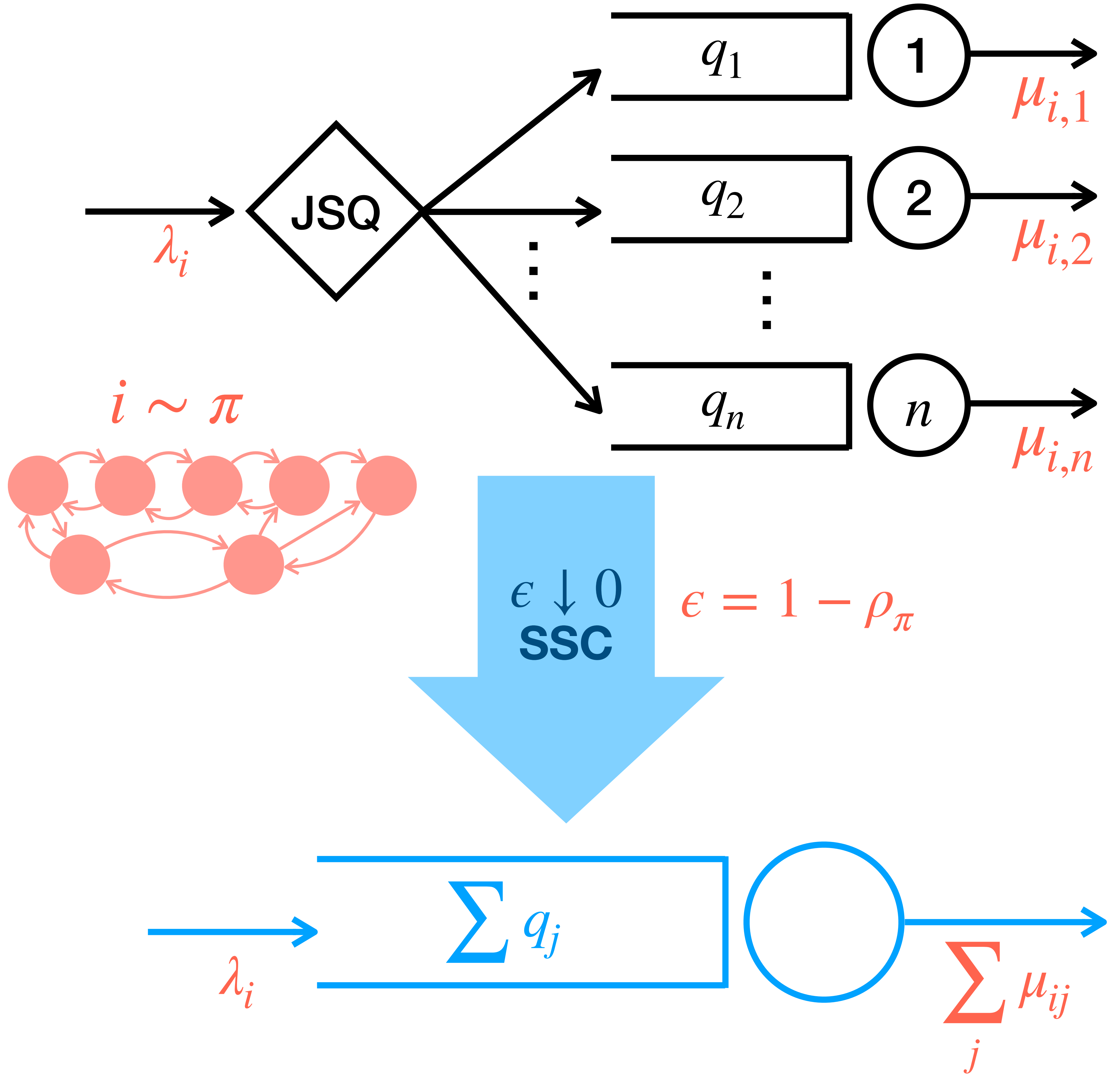
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$



State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

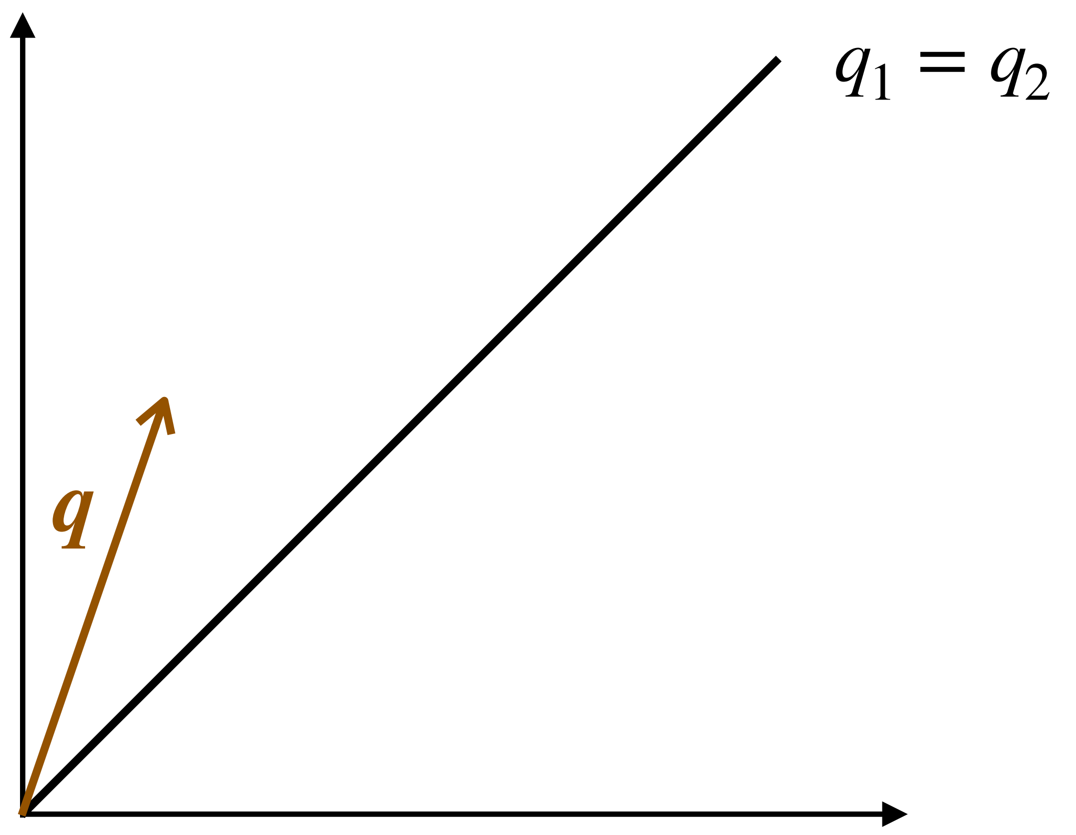


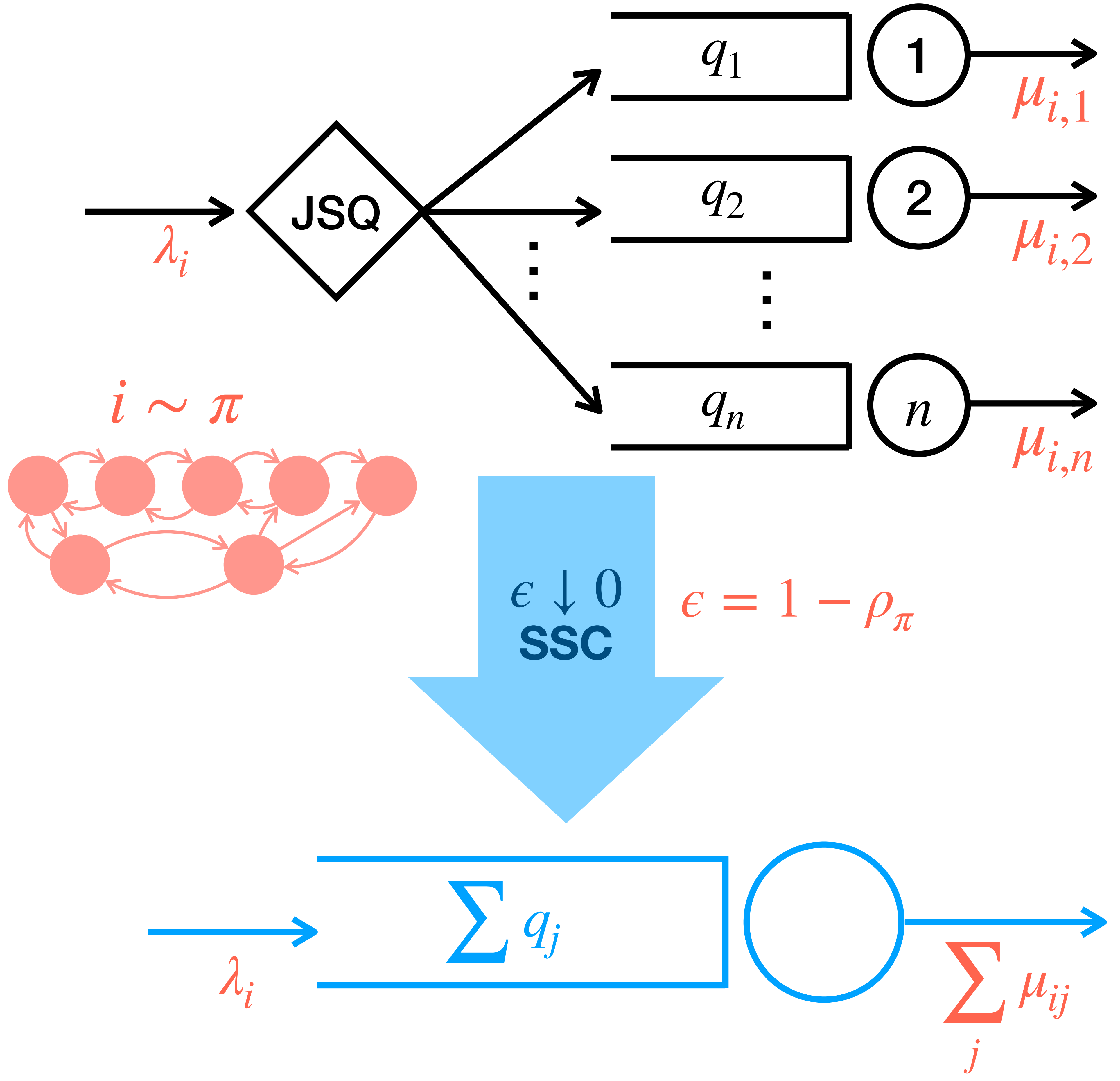


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

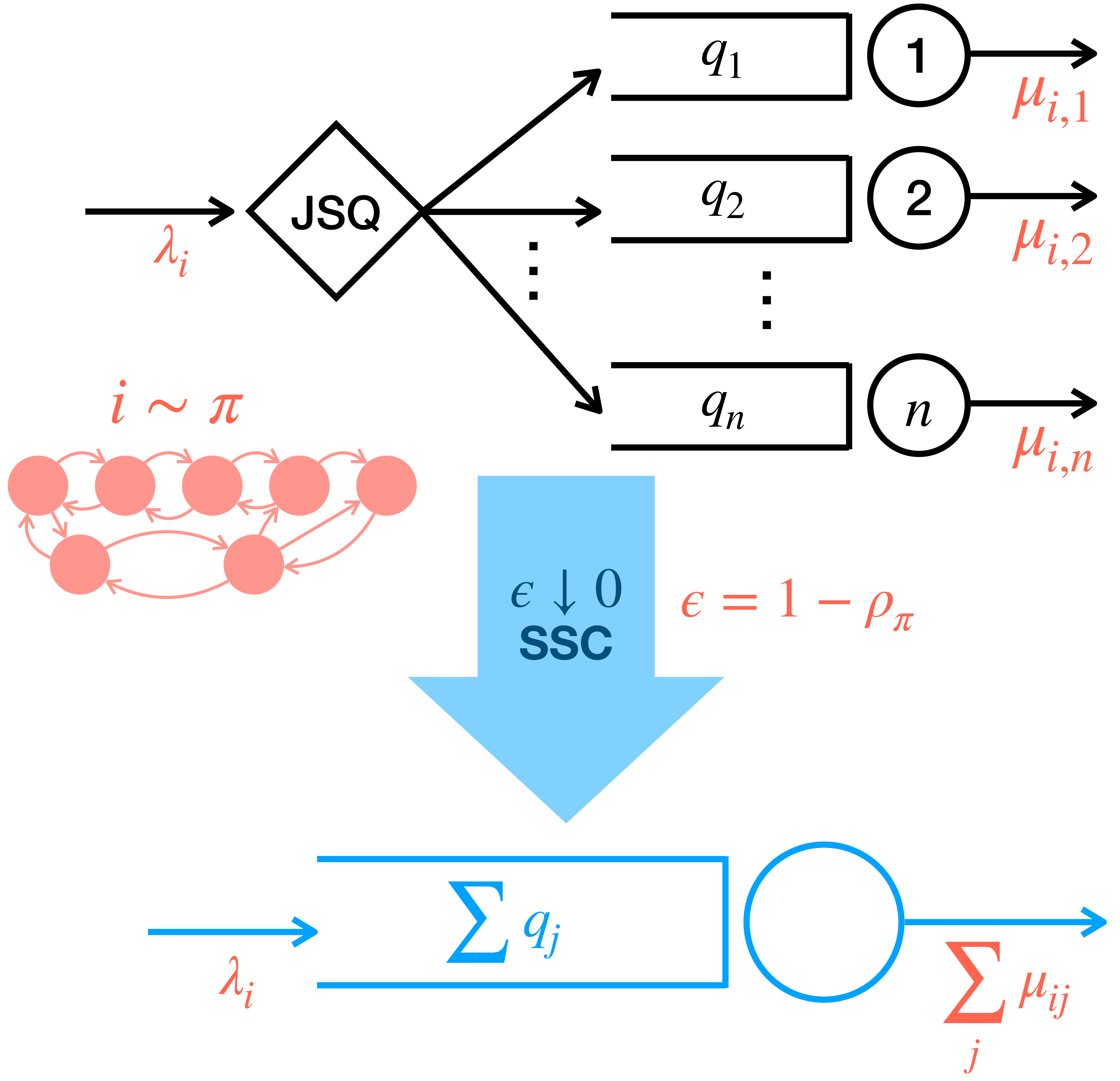


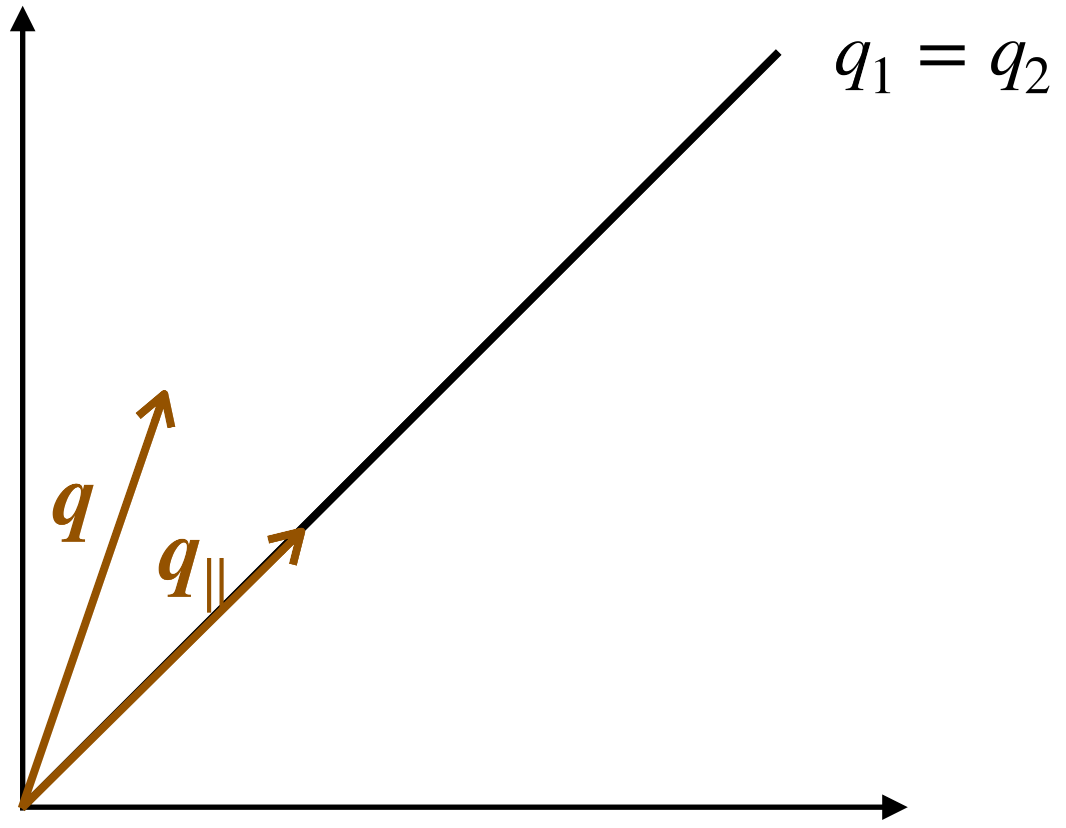


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

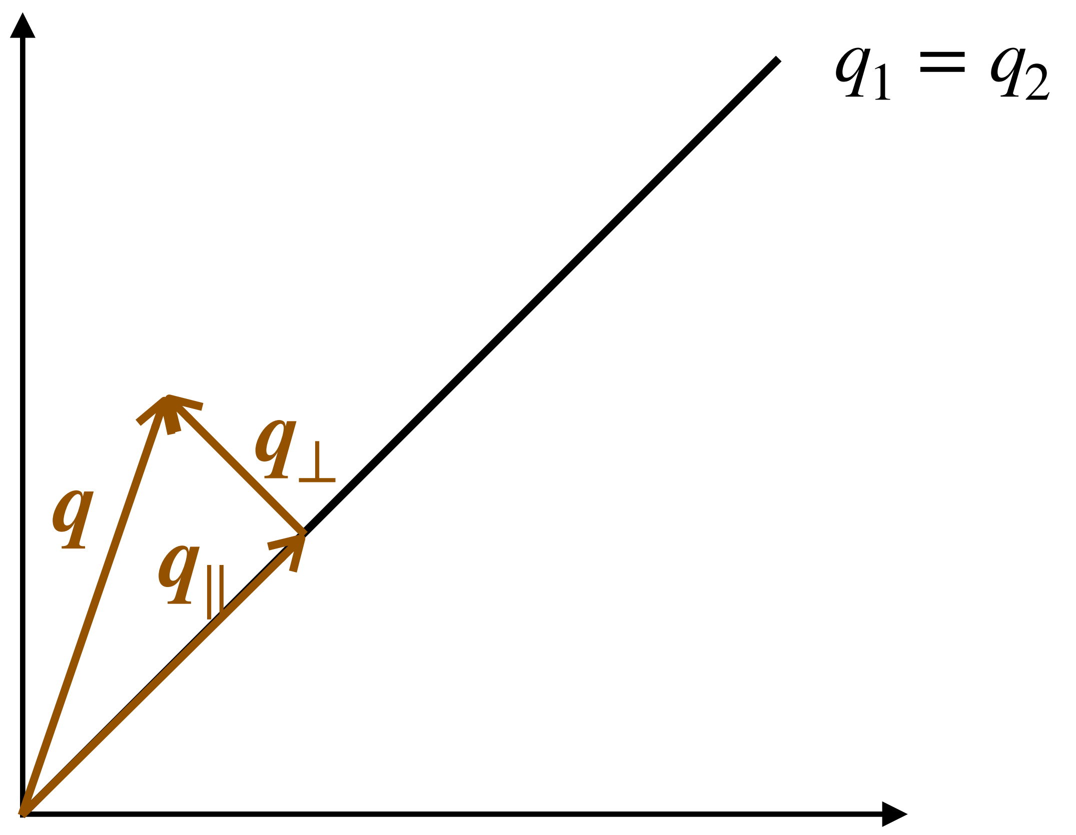
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

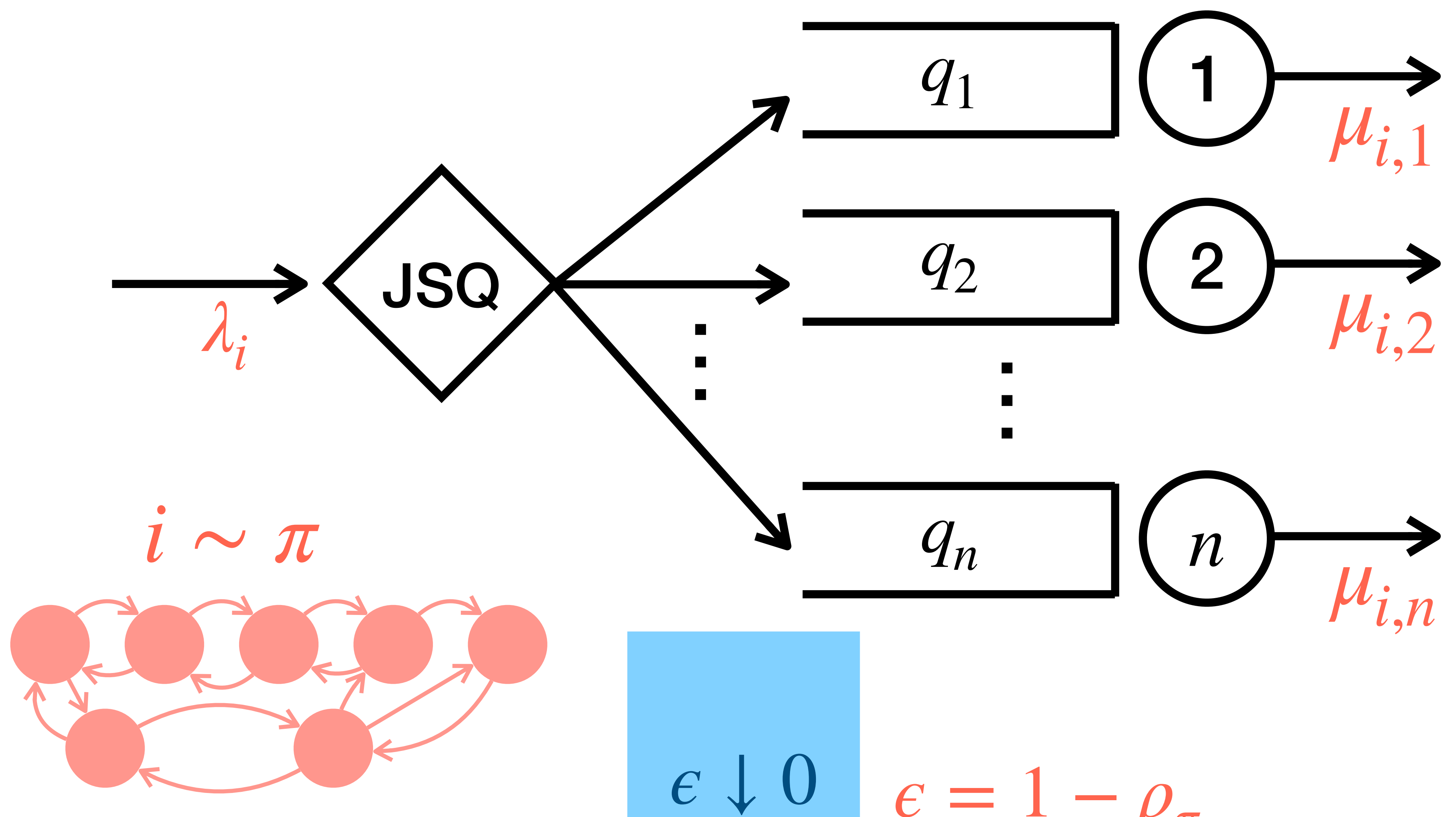


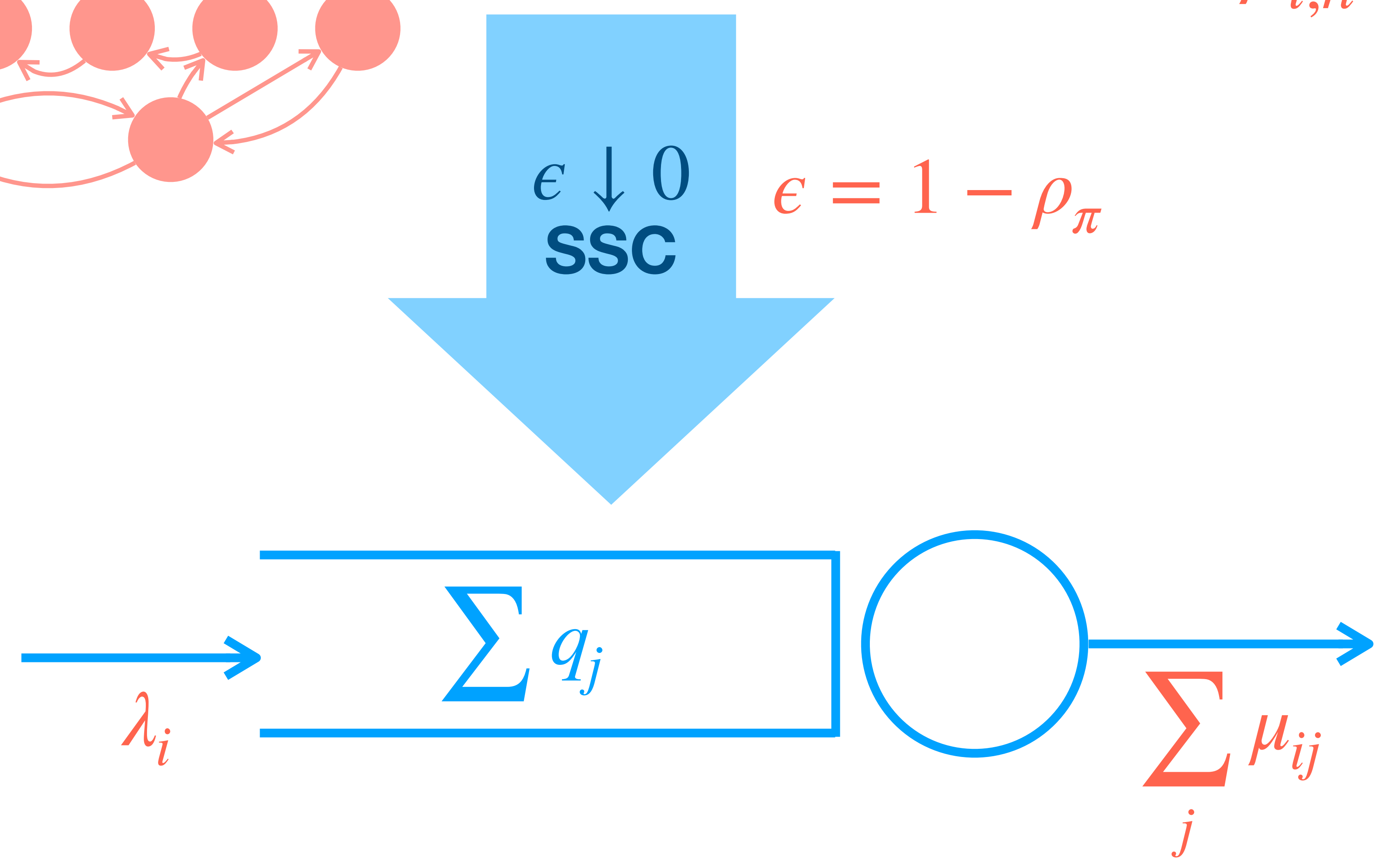


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

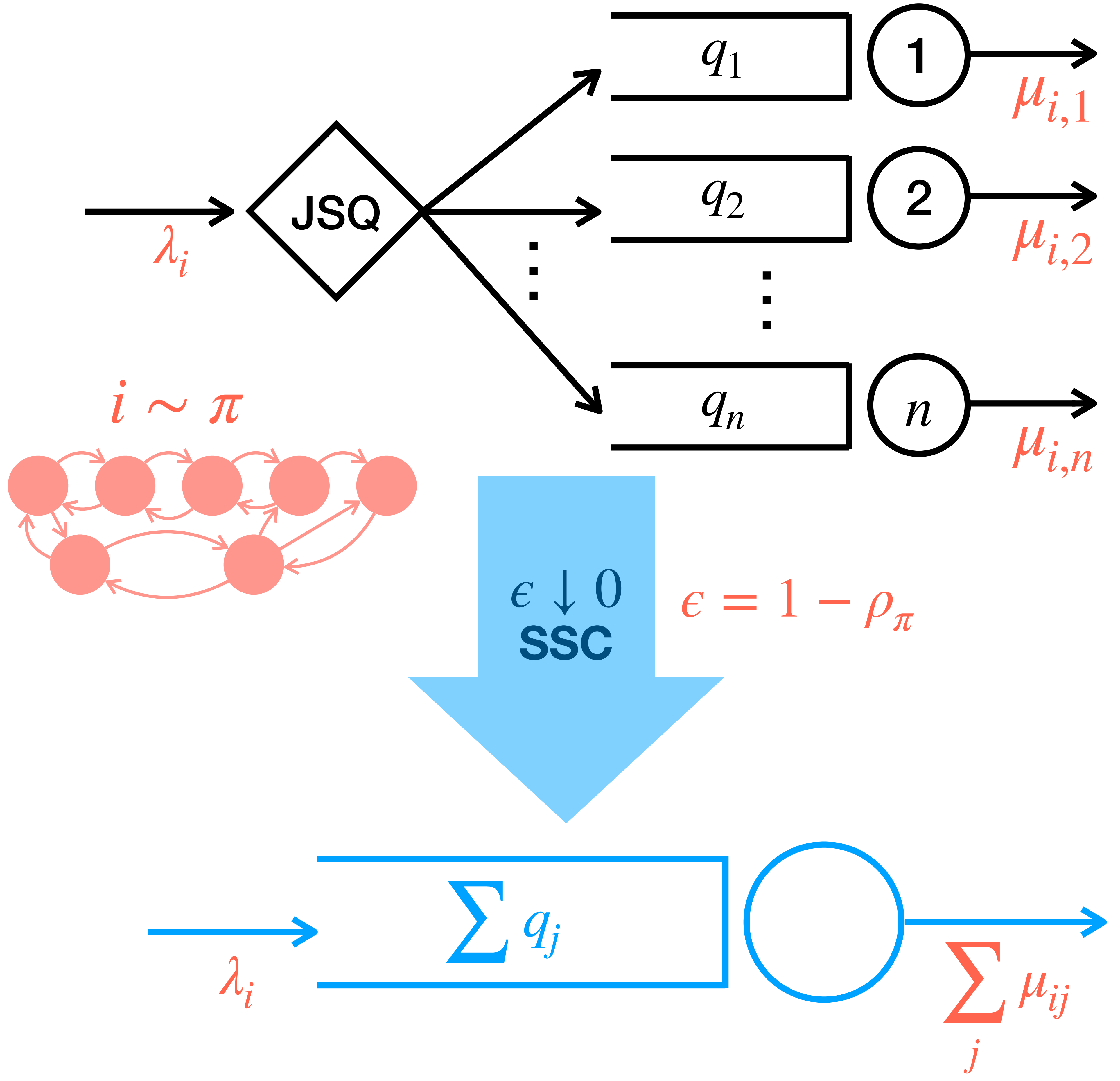


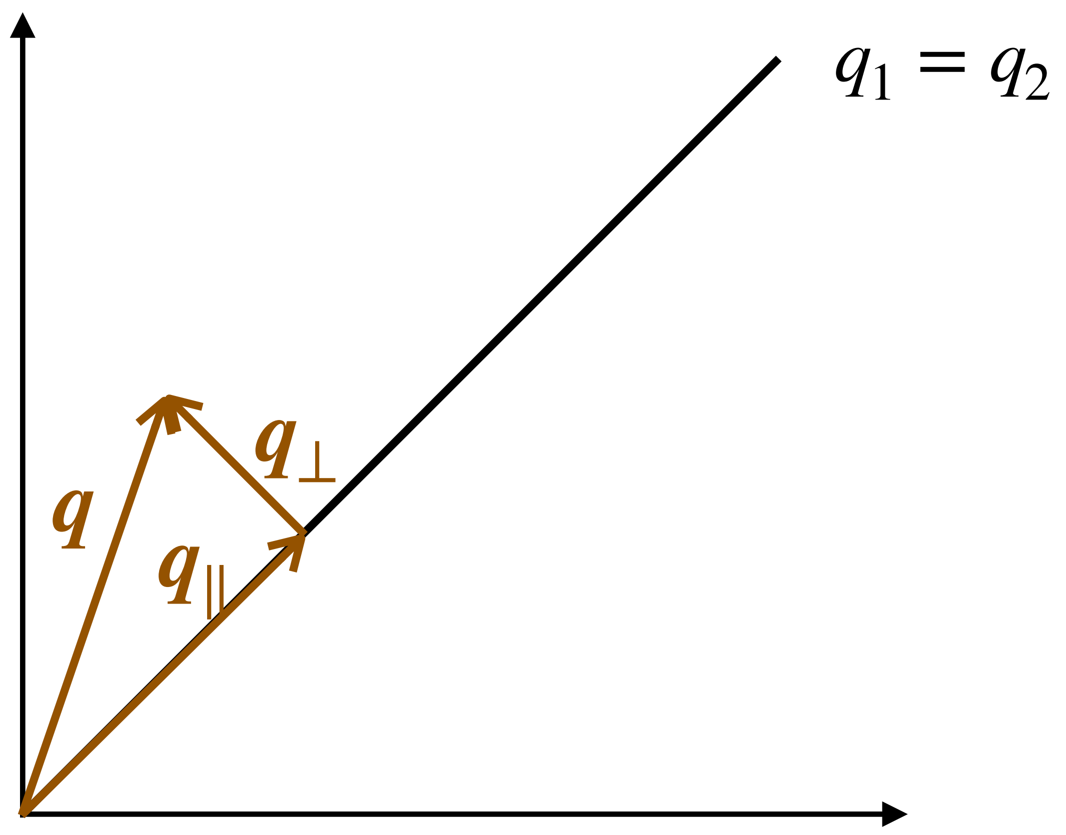


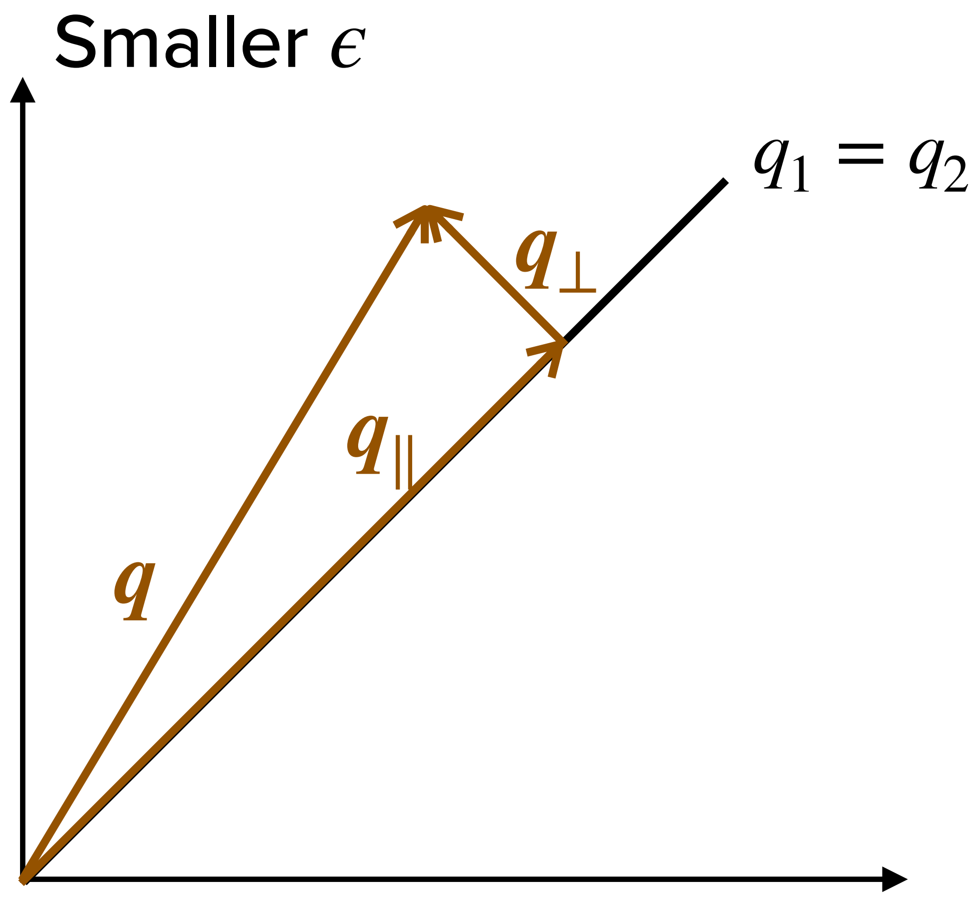


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

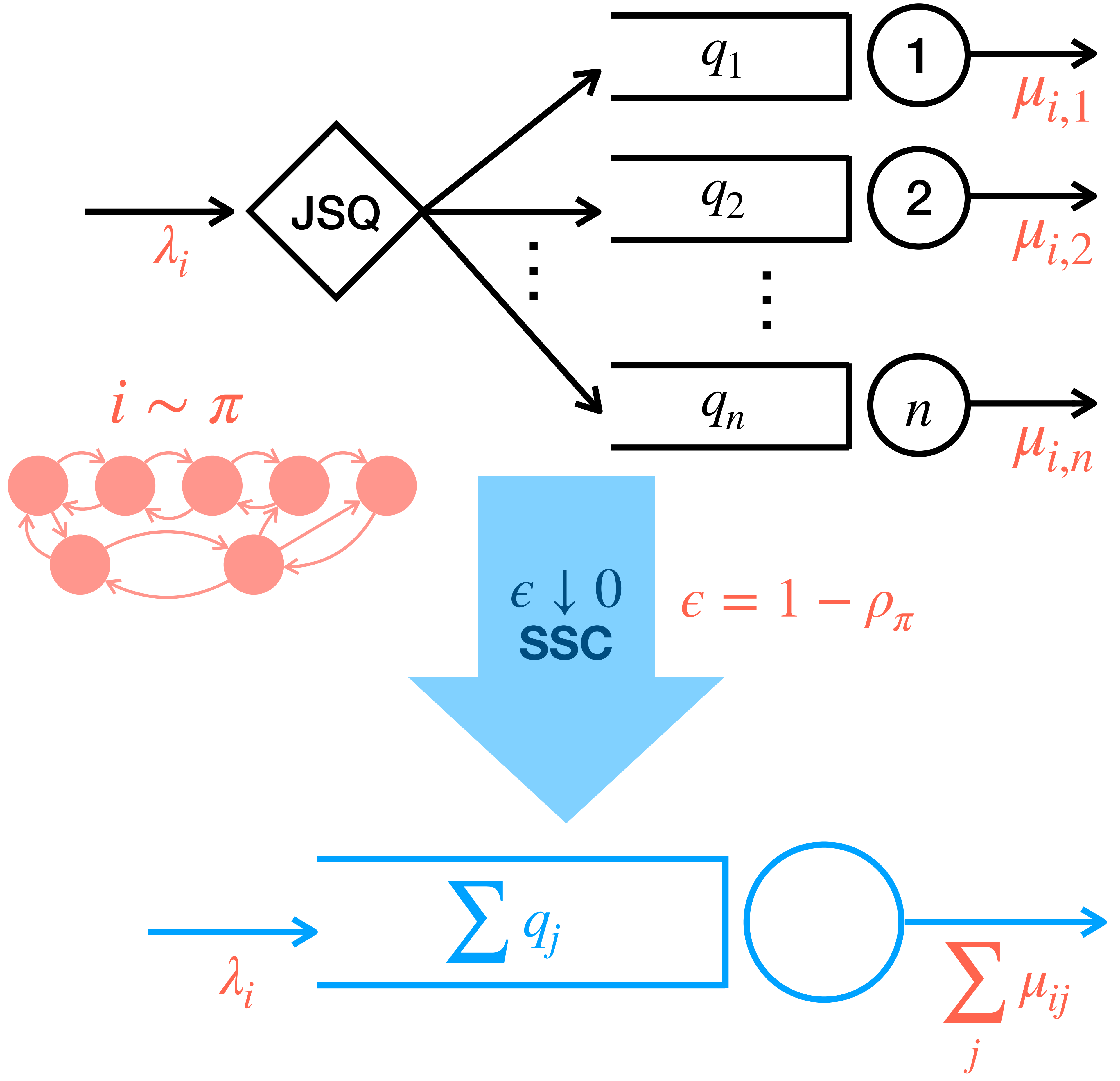


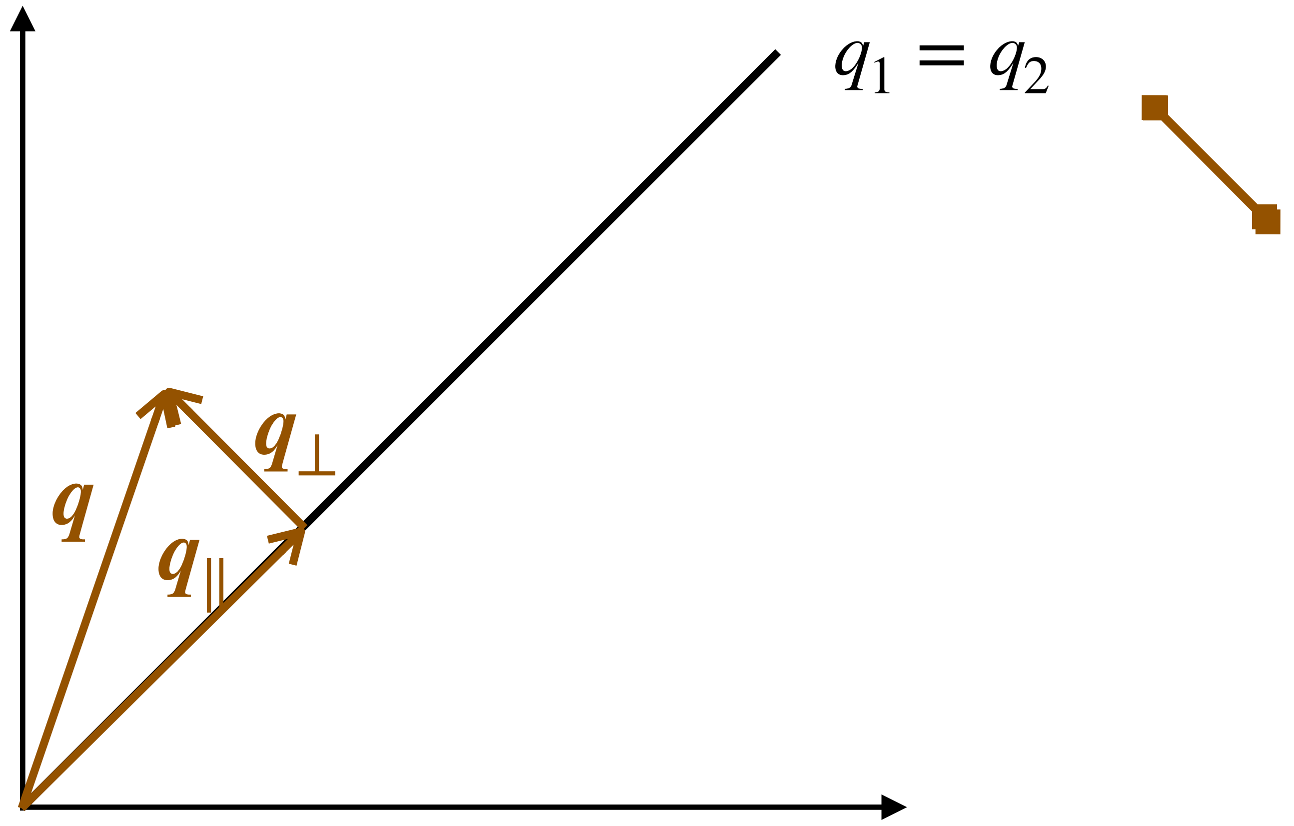


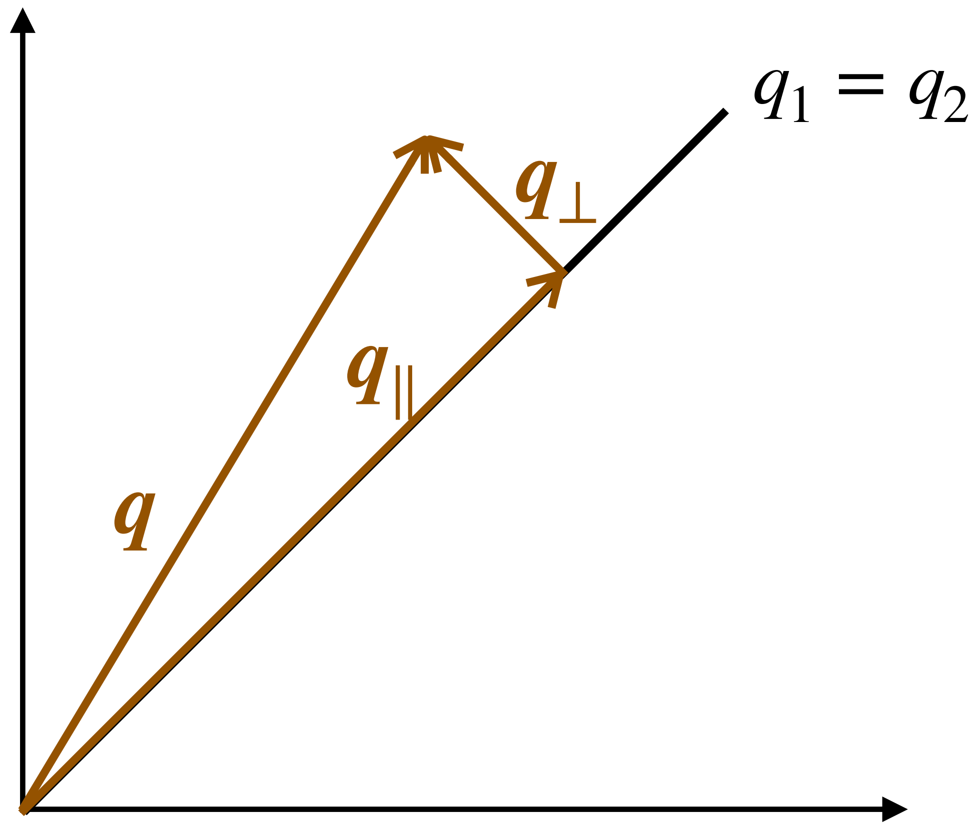


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \le M_m$$

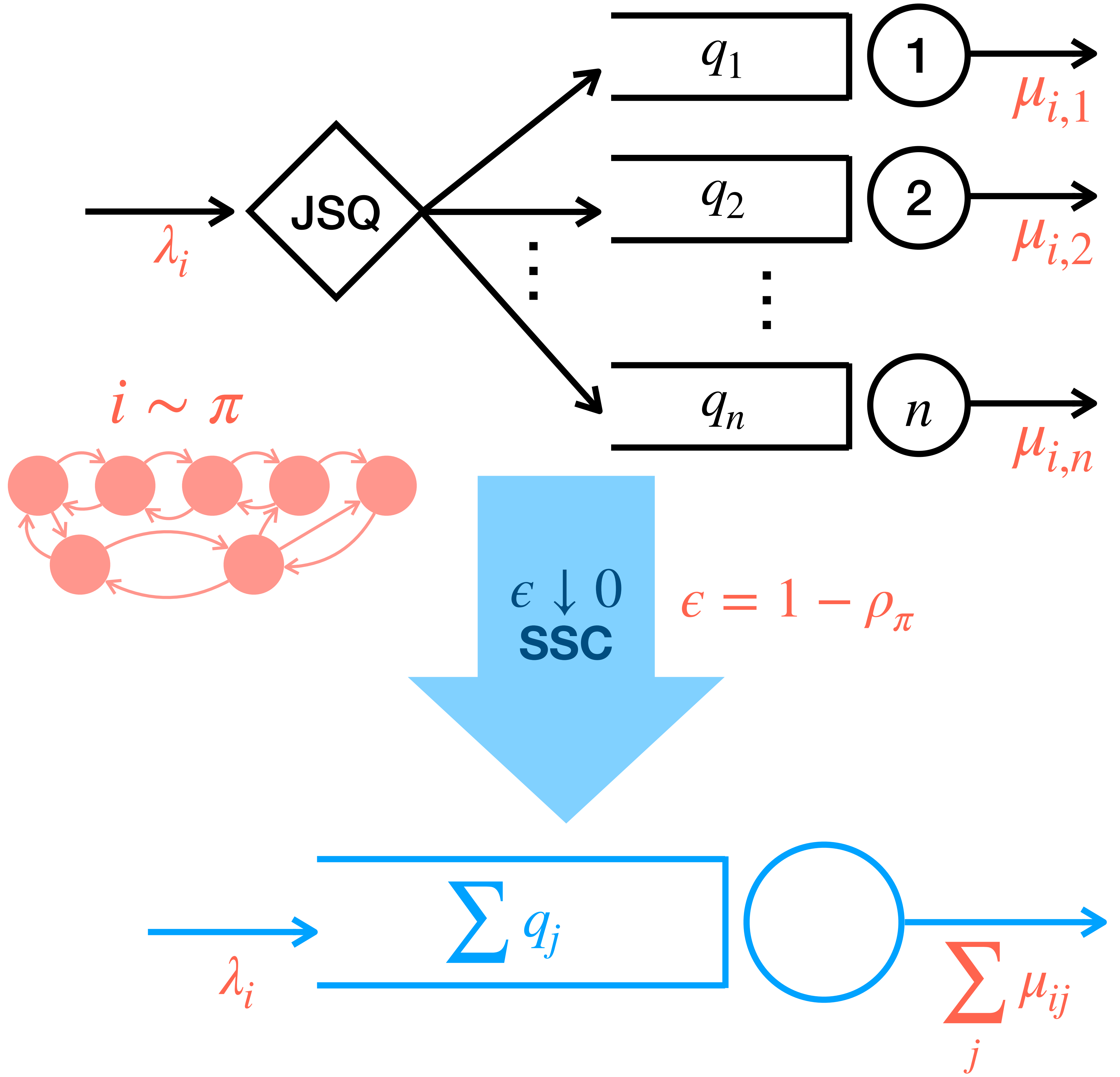


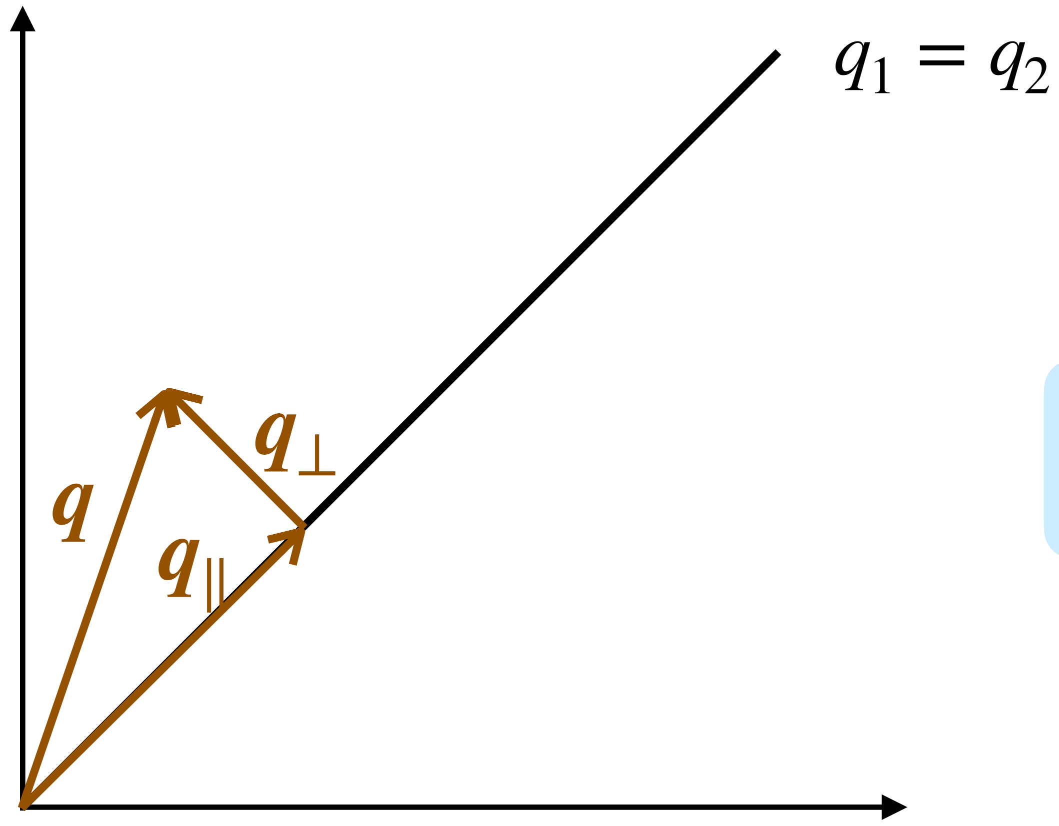




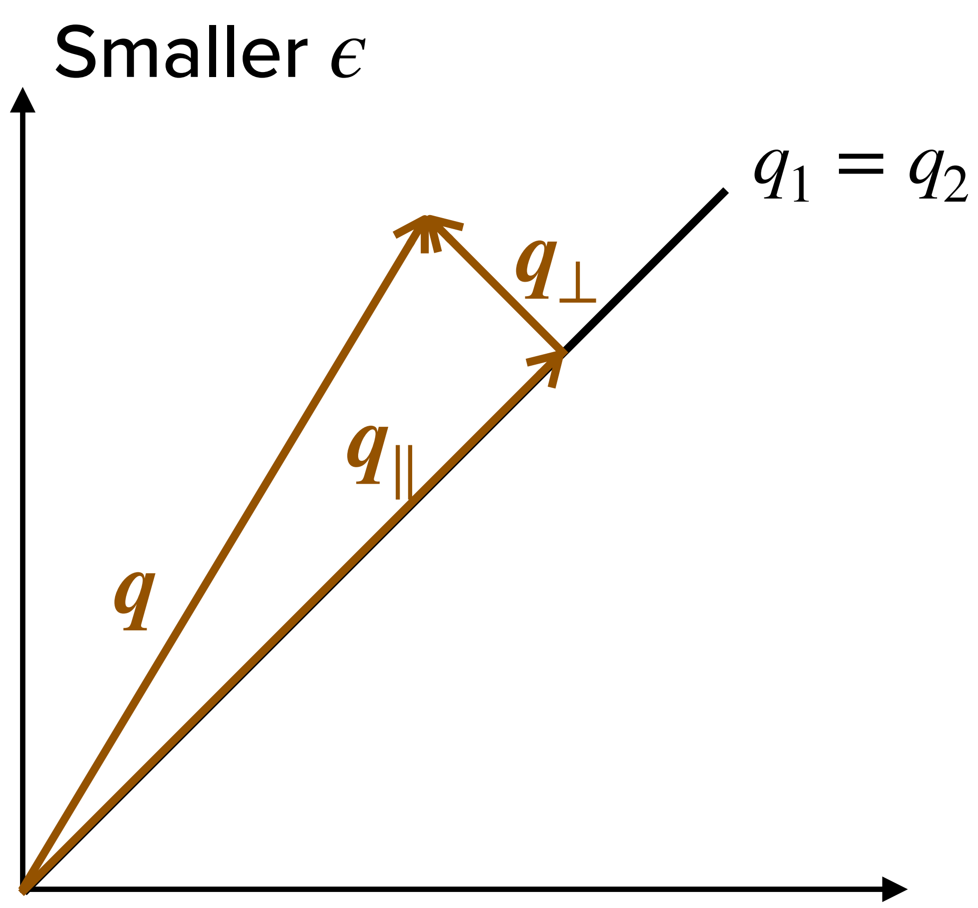
State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$





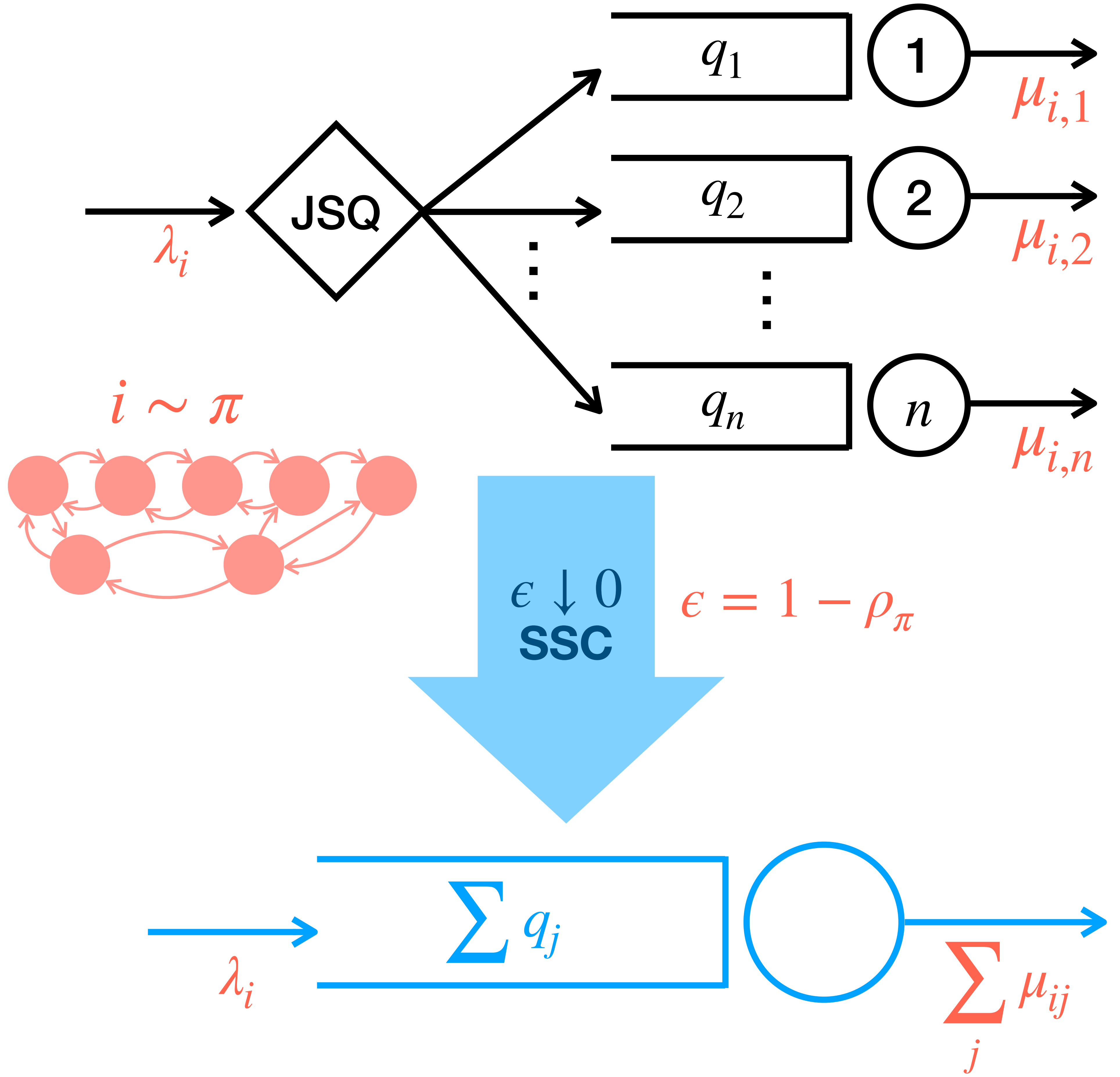
$\boldsymbol{q} \approx \boldsymbol{q}_{\|}$

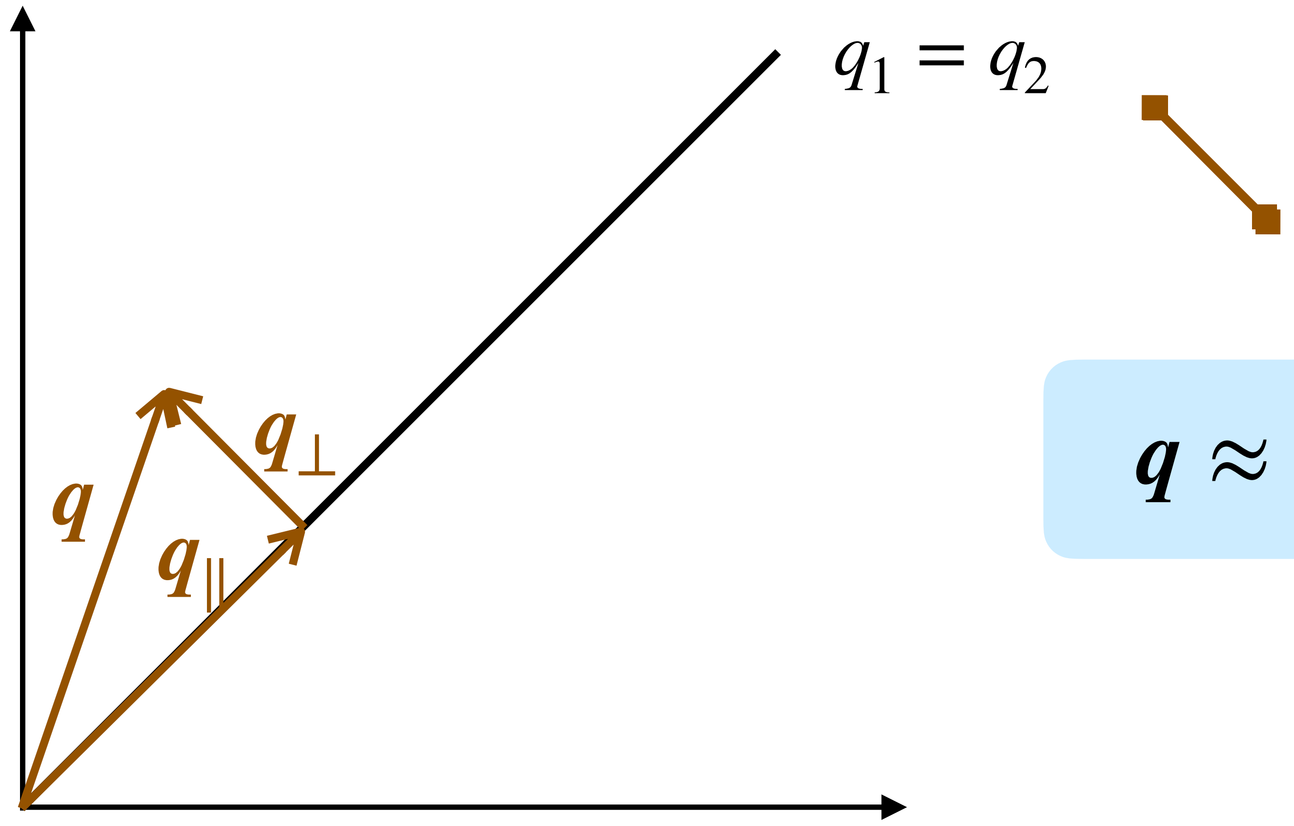


State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \le M_m$$

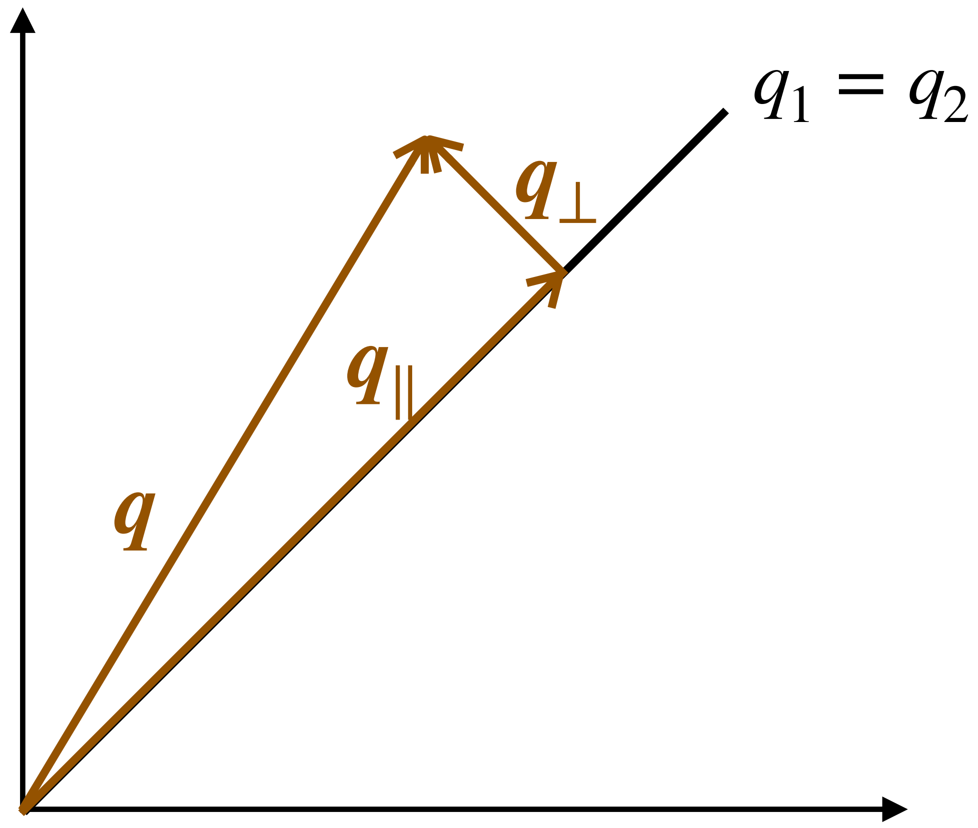
Proof: Drift analysis





$\boldsymbol{q} \approx \boldsymbol{q}_{\parallel}$

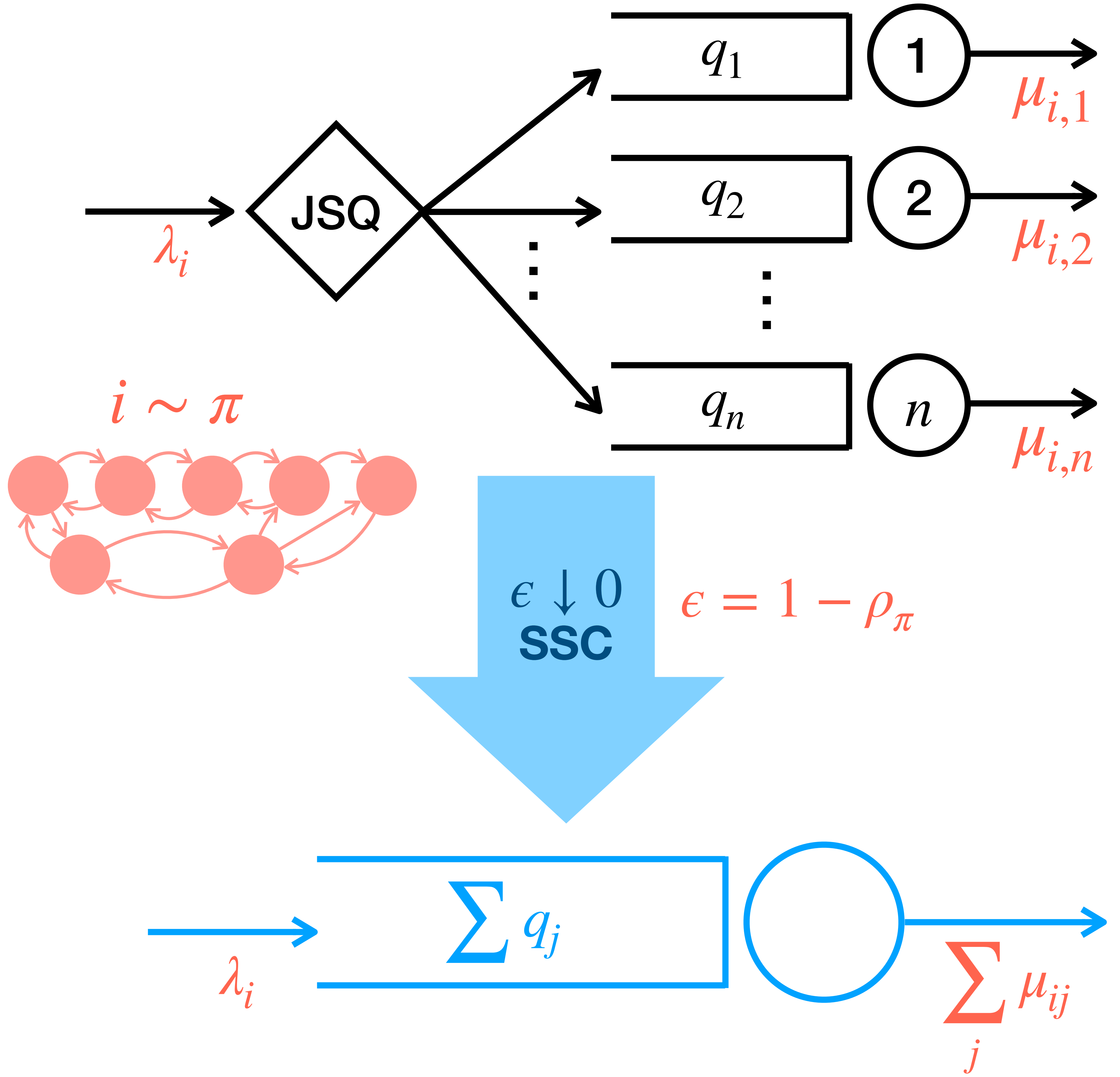
Smaller ϵ



State Space Collapse (SSC)

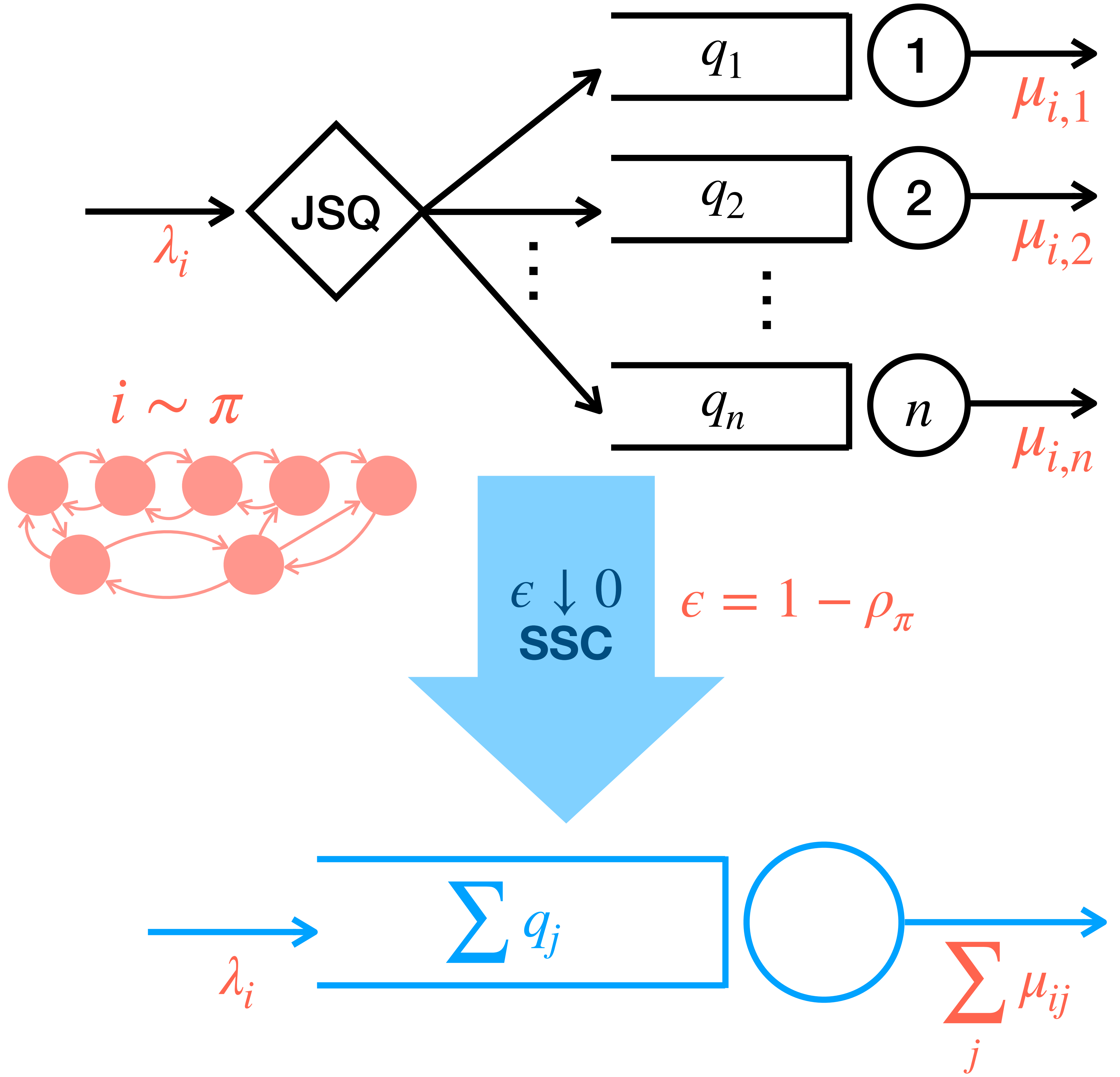
Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$



State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

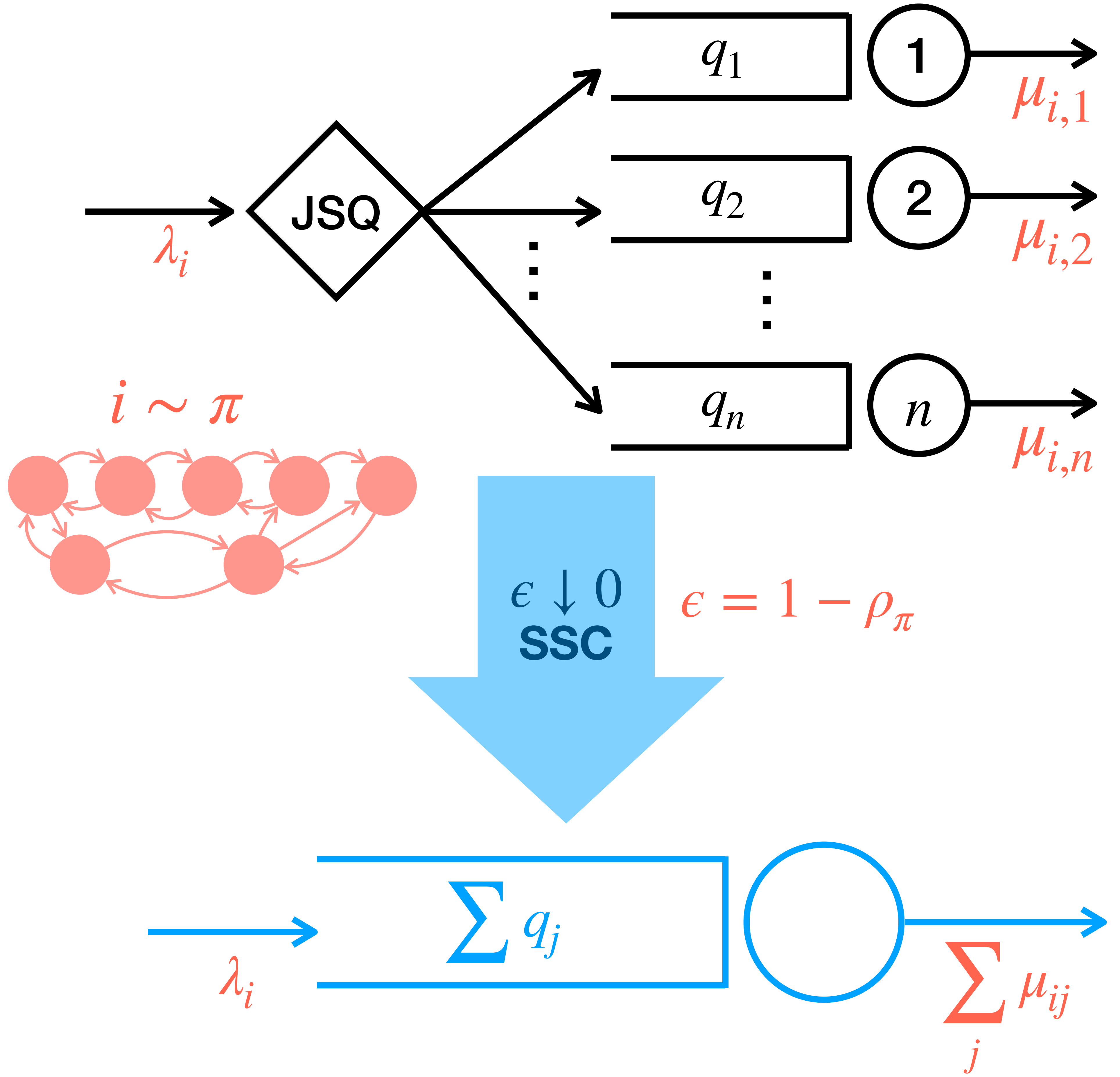


What is λ_i large enough?

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

State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\parallel}$:= projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\parallel}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$



✓ For homogeneous servers, $\lambda_i > 0$ is enough

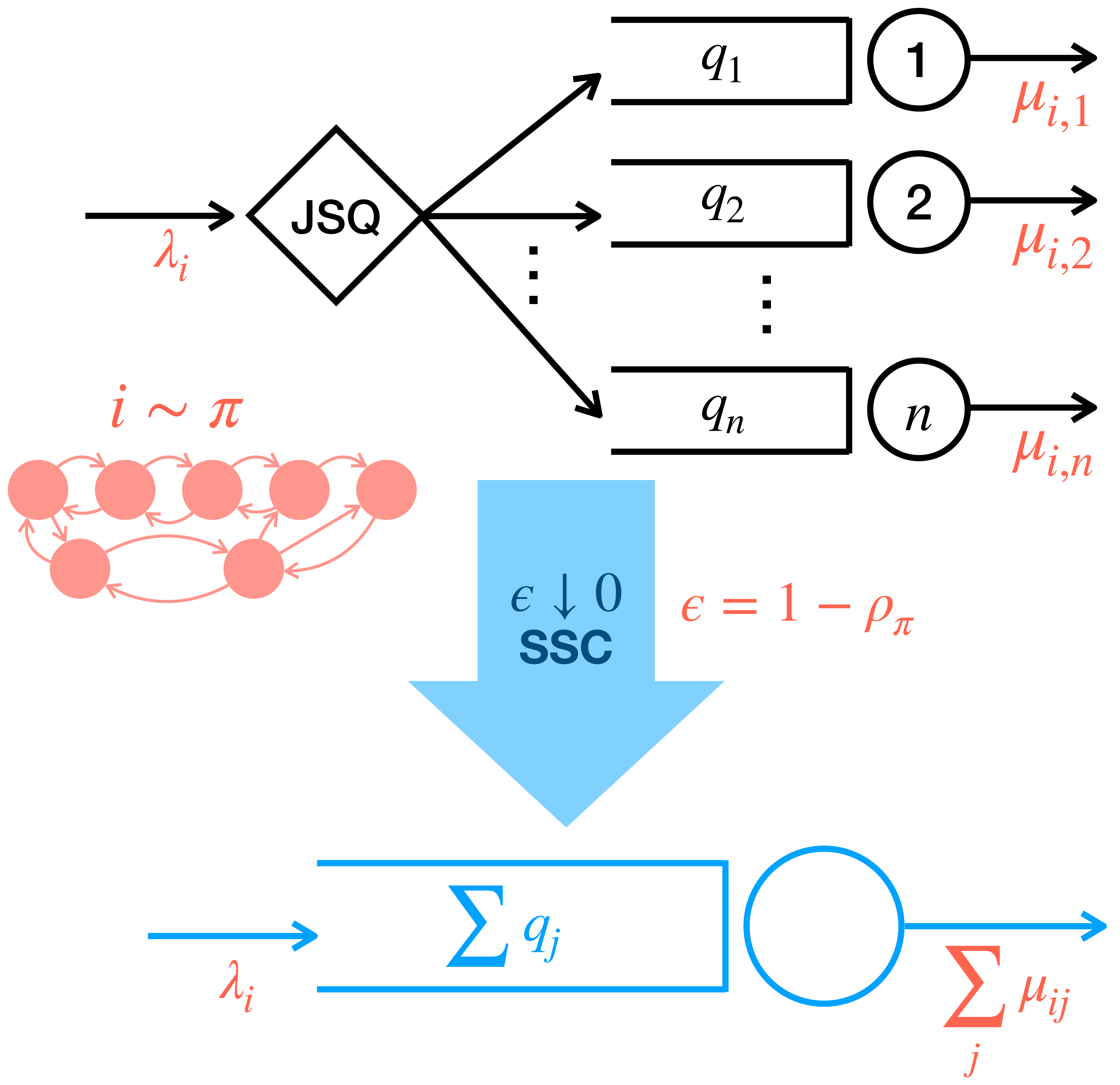
What is λ_i large enough?

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

State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$



What is λ_i large enough?

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

✓ For homogeneous servers, $\lambda_i > 0$ is enough

✓ "Enough" depends on heterogeneity of servers

State Space Collapse (SSC)

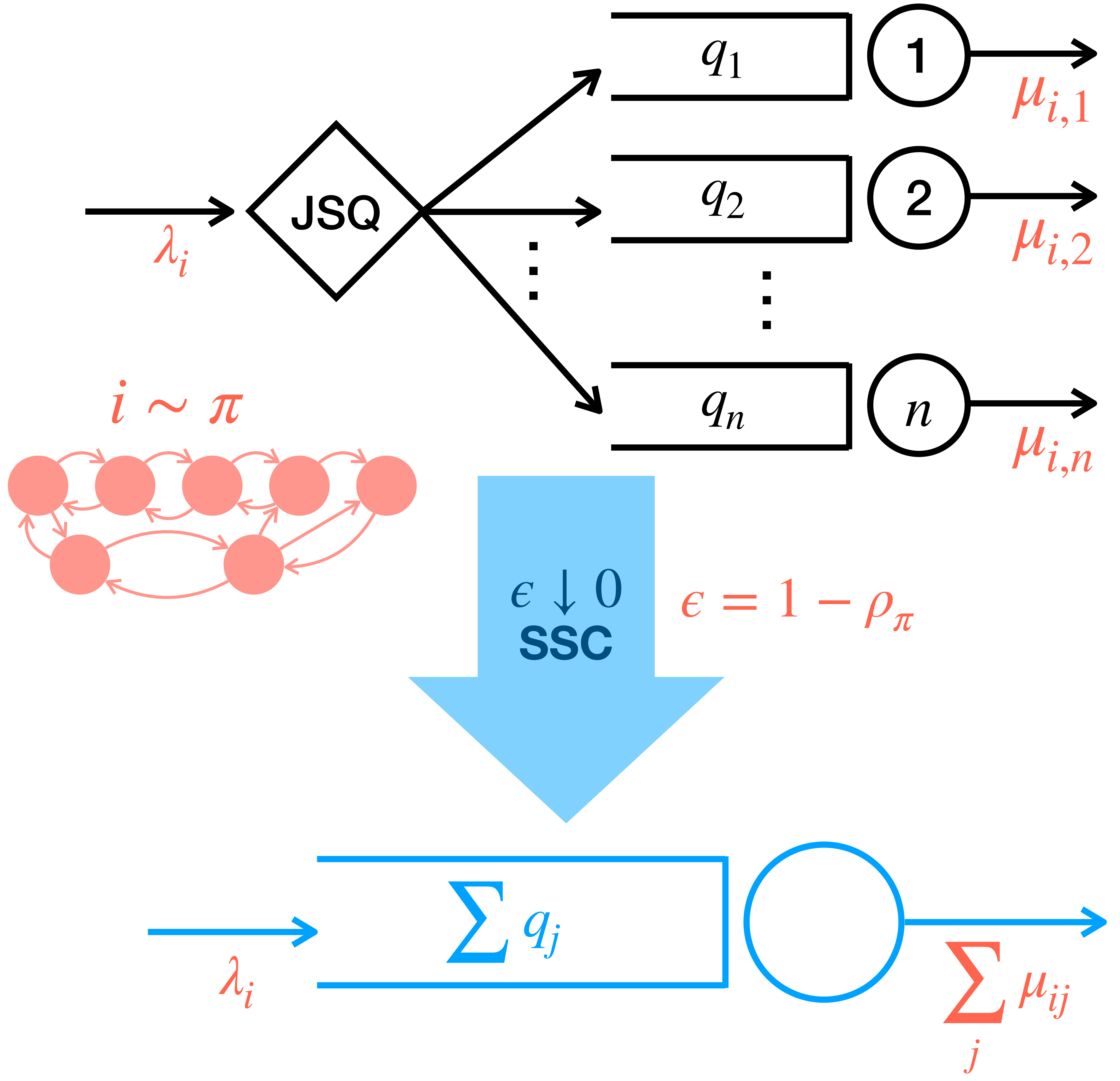
Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

What is λ_i large enough?

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

- ✓ For homogeneous servers, $\lambda_i > 0$ is enough
- ✓ "Enough" depends on heterogeneity of servers
- ✓ Arrivals need to cover range of service rates



State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,

$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \leq M_m$$

What is λ_i large enough?

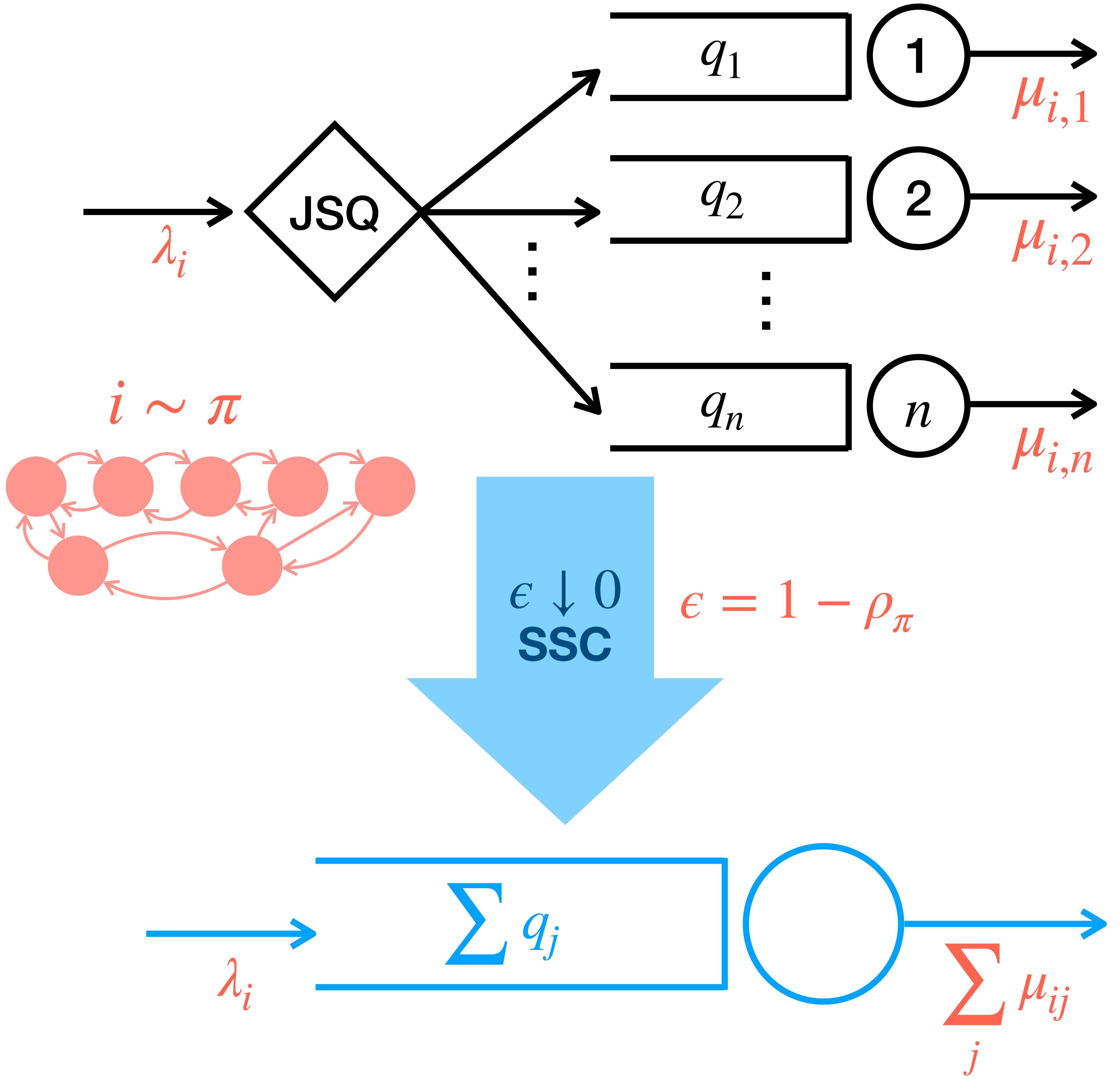
$$\lambda_i > \mu_{i\Sigma} - n \min_j \mu_{ij}$$

✓ For homogeneous servers, $\lambda_i > 0$ is enough

✓ "Enough" depends on heterogeneity of servers

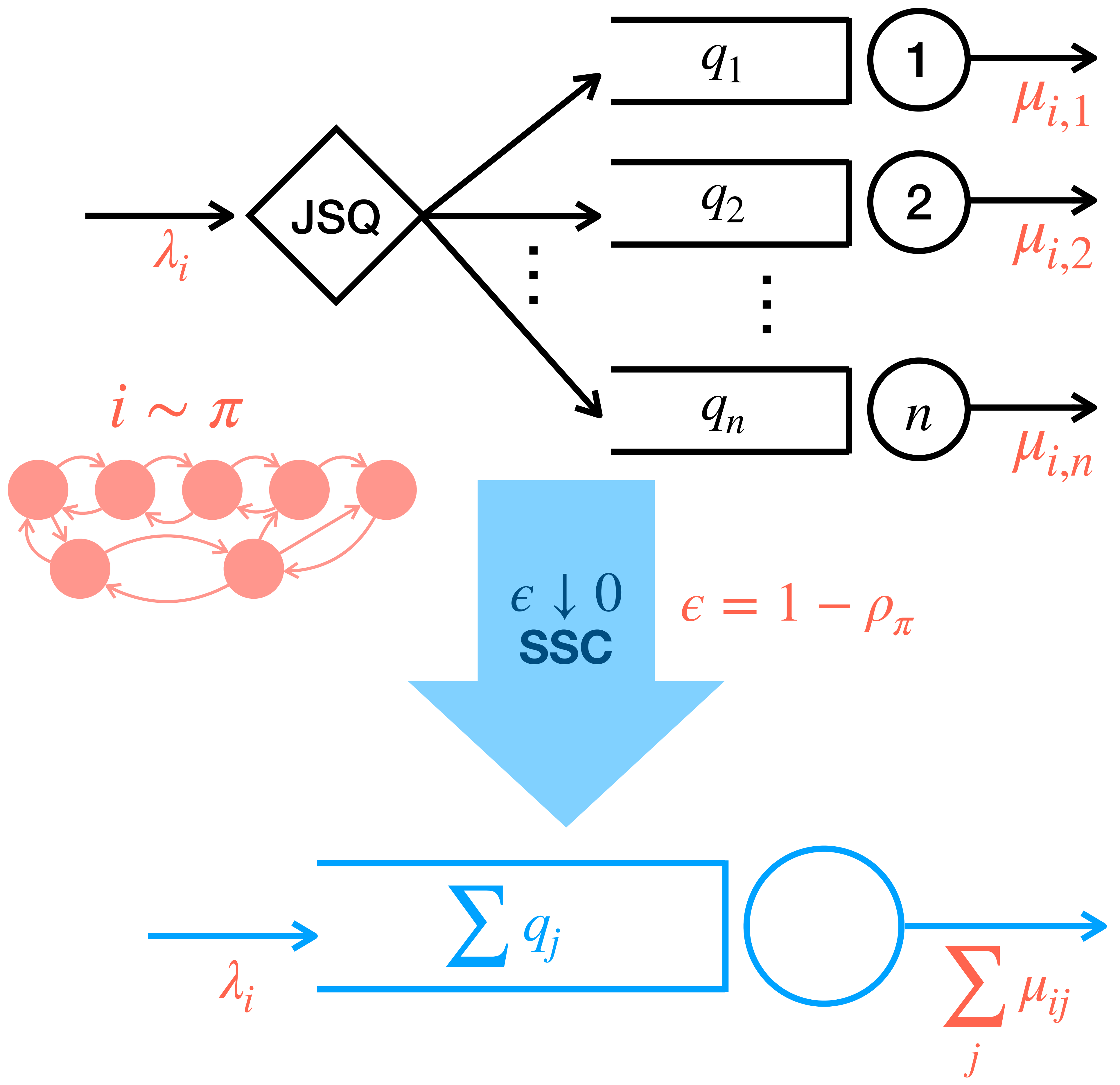
✓ Arrivals need to cover range of service rates

✓ JSQ only balances $\boldsymbol{q}$ with *enough arrivals*



State Space Collapse (SSC)

Proposition [HL, Grosof '25]:
$\boldsymbol{q}_{\|} :=$ projection of $\boldsymbol{q}$ on $\mathbf{1}$
$\boldsymbol{q}_{\perp} := \boldsymbol{q} - \boldsymbol{q}_{\|}$
If $\lambda_i > 0$ is large enough for each i, for each $m \in \mathbb{N}$,
$$\mathbb{E}\left[\|\boldsymbol{q}_{\perp}\|^m\right] \le M_m$$



What is λ_i large enough?

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

- ✓ 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 $\boldsymbol{q}$ with *enough arrivals*

When does JSQ balance the queue lengths?

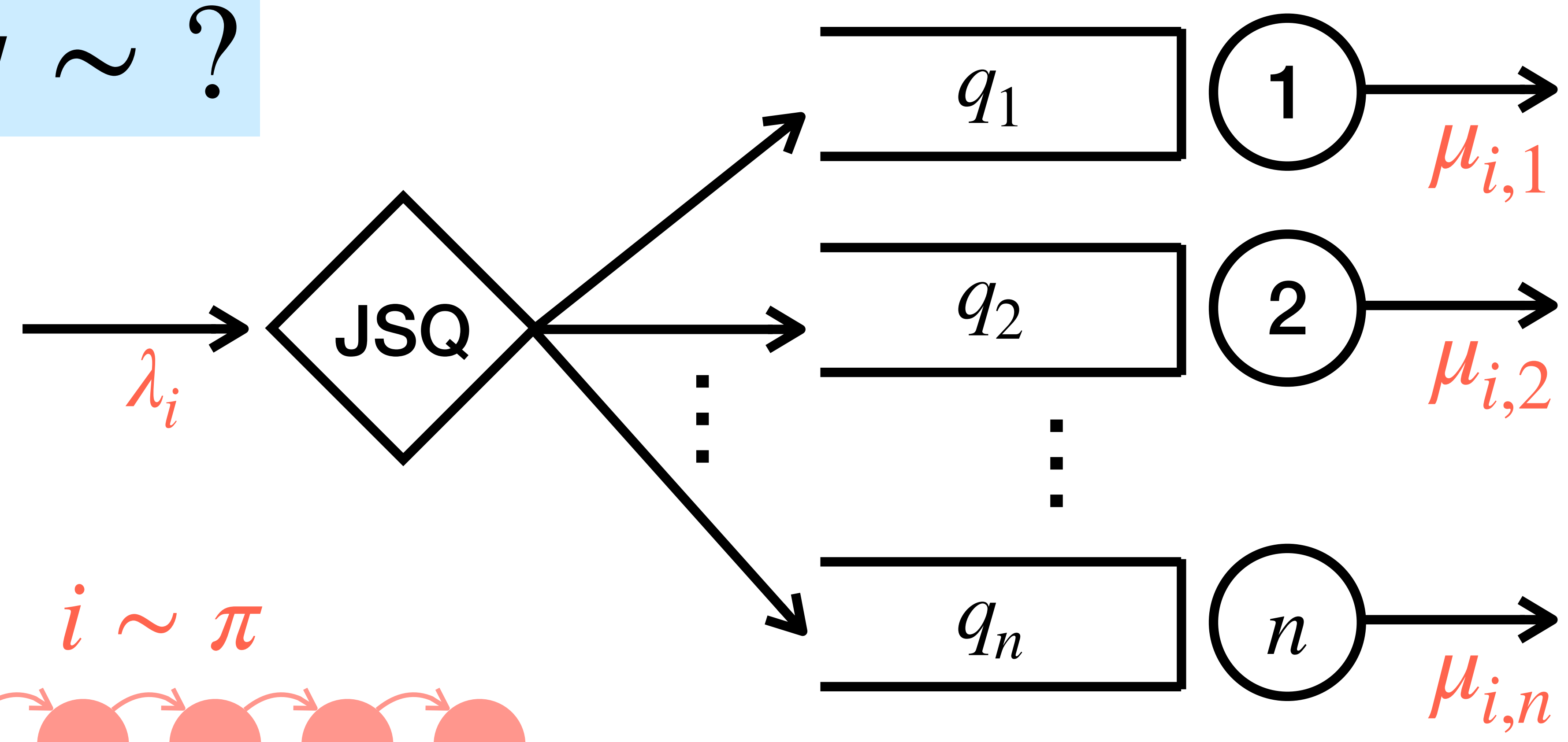
Proof Sketch

Theorem [HL, Grosof '25]:
If $\lambda_i > 0$ is large enough for each i, then as $\epsilon \downarrow 0$

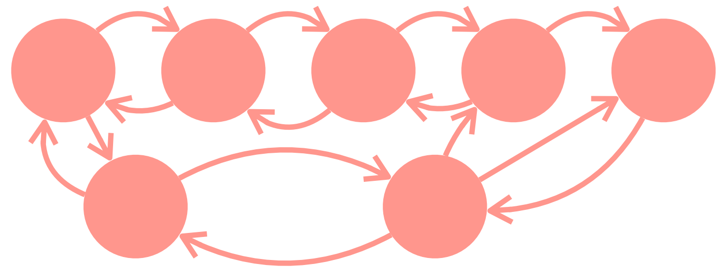
$$\epsilon\, \boldsymbol{q} \Longrightarrow Exp\left(\text{Mean} = 1 + \frac{k^*}{\mu_\Sigma}\right)$$

≈ Variance of arrival and service times in steady state

$\epsilon\, \boldsymbol{q} \sim ?$

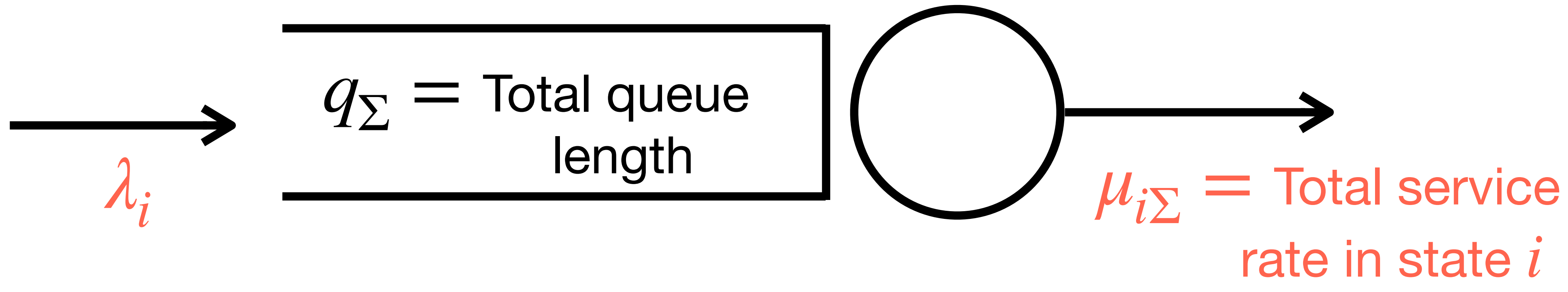


$i \sim \pi$



Step 1: State Space Collapse (SSC)

$\boldsymbol{q} \approx \left(\frac{q_\Sigma}{n}\right) \mathbf{1}$, 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

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 $1-\rho$

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 $1-\rho$

$$\epsilon q^{(\epsilon)}(t)$$

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 $1-\rho$

$$\epsilon q^{(\epsilon)}(t) \xrightarrow{t \to \infty} \epsilon q^{(\epsilon)}(\infty)$$

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 $1 - \rho$

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \\
\tilde{q}(t) &
\end{tikzcd}$$

RBM

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 $1-\rho$

$$\begin{tikzcd} \epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \\ \widetilde{q}(t) \arrow[r, "t \to \infty"] & q^* \end{tikzcd}$$

RBM

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 $1 - \rho$

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\
\widetilde{q}(t) \arrow[r, "t \to \infty"] & q^*
\end{tikzcd}$$

Interchange of limits

RBM

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 $1-\rho$

$$\begin{tikzcd} \epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\ \widetilde{q}(t) \arrow[r, "t \to \infty"] & q^* \end{tikzcd}$$

Interchange of limits

RBM

Direct Methods

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 $1-\rho$

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\
\widetilde{q}(t) \arrow[r, "t \to \infty"] & q^*
\end{tikzcd}$$

Interchange of limits

RBM

Direct Methods

- **Stein's method**
  Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*

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 $1 - \rho$

$$\begin{tikzcd} \epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\ \widetilde{q}(t) \arrow[r, "t \to \infty"] & q^* \end{tikzcd}$$

Interchange of limits

RBM

Direct Methods

- **Stein's method**
  Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*
- **Drift Method**
  Inductively compute $\mathbb{E}\left[(\epsilon q)^k\right]$ for $k \in \mathbb{N}$

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 $1-\rho$

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\
\widetilde{q}(t) \arrow[r, "t \to \infty"] & q^*
\end{tikzcd}$$

Interchange of limits

RBM

Direct Methods

- **Stein's method**
  Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*
- **Drift Method**
  Inductively compute $\mathbb{E}\left[(\epsilon q)^k\right]$ for $k \in \mathbb{N}$
- **BAR approach**
  Analyze $e^{\epsilon q(\infty)}$ and bound jumps appropriately

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 $1-\rho$

$$\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"', "\text{Interchange of limits}"] \\
\widetilde{q}(t) \arrow[r, "t \to \infty"] & q^*
\end{tikzcd}$$

RBM

Direct Methods

- **Stein's method**
  Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*
- **Drift Method**
  Inductively compute $\mathbb{E}\left[(\epsilon q)^k\right]$ 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 $\mathbb{E}\left[e^{\epsilon q^{(\epsilon)}(\infty)}\right]$

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 $1 - \rho$

$$
\begin{tikzcd}
\epsilon q^{(\epsilon)}(t) \arrow[r, "t \to \infty"] \arrow[d, "\epsilon \downarrow 0"'] & \epsilon q^{(\epsilon)}(\infty) \arrow[d, "\epsilon \downarrow 0"'] \\
\tilde{q}(t) \arrow[r, "t \to \infty"] & q^*
\end{tikzcd}
$$

Interchange of limits

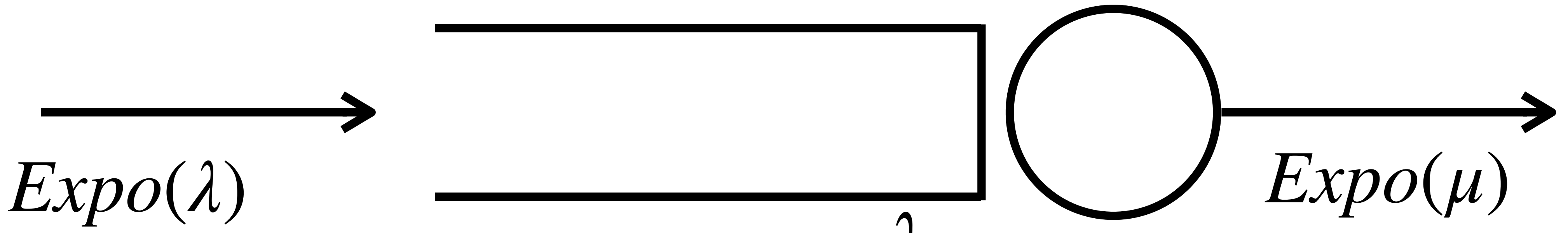
RBM

Direct Methods

- **Stein's method**
  Bound Wasserstein distance between $\epsilon q^{(\epsilon)}(\infty)$ and q^*
- **Drift Method**
  Inductively compute $\mathbb{E}\left[(\epsilon q)^k\right]$ 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 $\mathbb{E}\left[e^{\epsilon q^{(\epsilon)}(\infty)}\right]$
  - ✓ Tractable analysis
  - ✓ Compute distribution
  - ✓ Obtain tail bounds

Transform Method for $M/M/1$ queue

Transform Method for $M/M/1$ queue

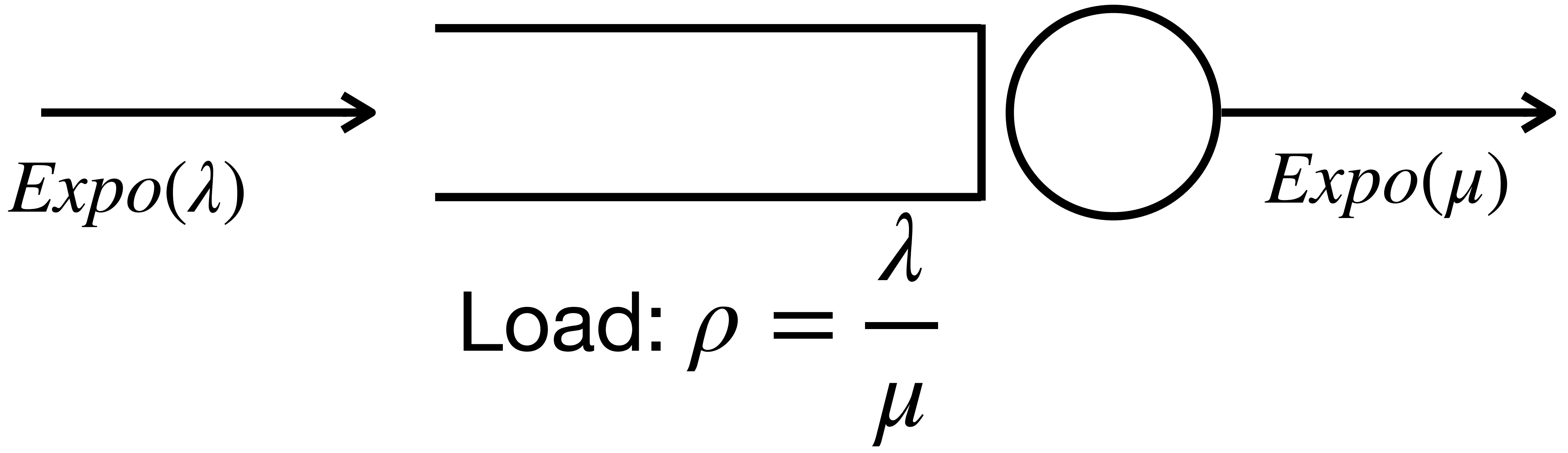


Load: $\rho = \dfrac{\lambda}{\mu}$

Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

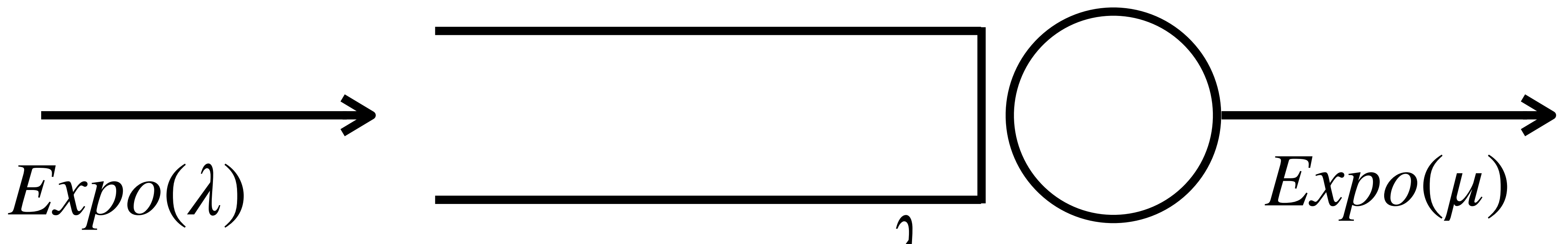
$\varphi_\theta(q) = e^{\theta q}$, $\theta \in \mathbb{R}$



Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

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



Load: $\rho = \dfrac{\lambda}{\mu}$

Drift:
$\Delta\varphi_\theta(q) = \mathbb{E}\left[e^{\theta q(t+)} - e^{\theta q} \,\middle|\, q(t) = q\right]$

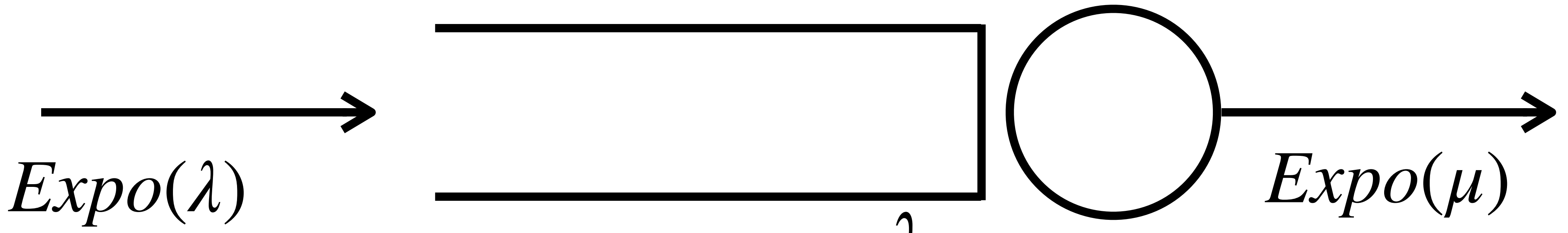
Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

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

Arrival / Departure

$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$



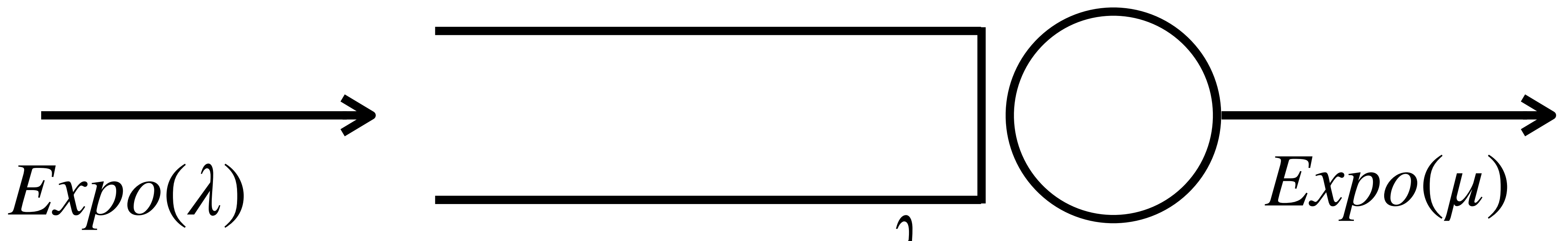
Load: $\rho = \dfrac{\lambda}{\mu}$

Drift:
$$\Delta\varphi_\theta(q) = \mathbb{E}\left[e^{\theta q(t+)} - e^{\theta q} \,\middle|\, q(t) = q\right]$$

Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

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



Load: $\rho = \dfrac{\lambda}{\mu}$

Arrival — Departure

$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$

$$= e^{\theta q}\left(e^{-\theta} - 1\right)\left(\mu - \lambda e^{\theta}\right) - \mu\left(e^{-\theta} - 1\right) e^{\theta q} 1_{\{q=0\}}$$

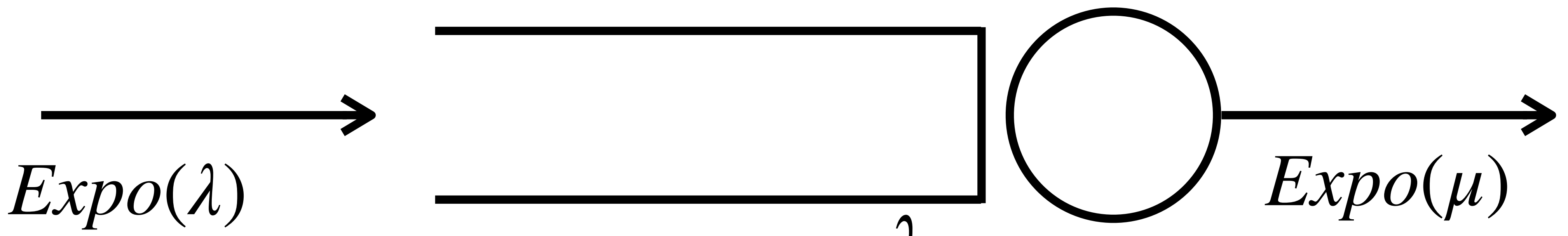
Drift:

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

Transform Method for $M/M/1$ queue

Step 1: Drift of exponential test function

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



Load: $\rho = \dfrac{\lambda}{\mu}$

Arrival   Departure

$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$

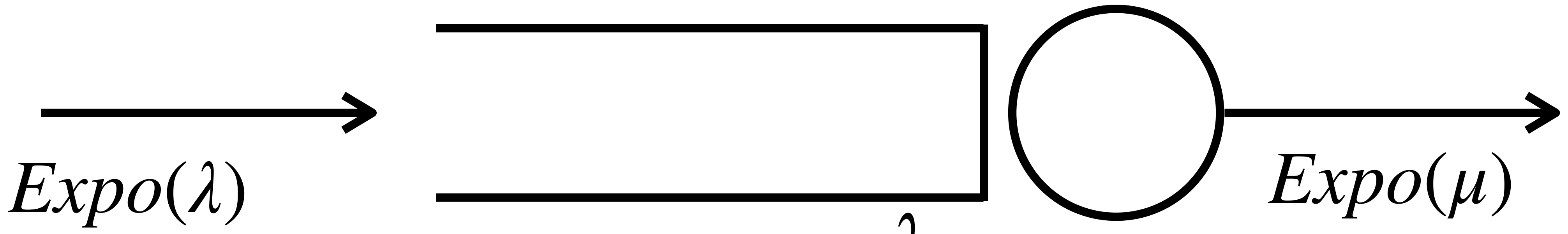
$$= e^{\theta q}\left(e^{-\theta} - 1\right)\left(\mu - \lambda e^{\theta}\right) - \mu\left(e^{-\theta} - 1\right) e^{\theta q} 1_{\{q=0\}}$$

$$= 1_{\{q=0\}}$$

Drift:

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

Transform Method for $M/M/1$ queue



Load: $\rho = \dfrac{\lambda}{\mu}$

Drift:
$\Delta\varphi_\theta(q) = \mathbb{E}\left[e^{\theta q(t+)} - e^{\theta q} \,\middle|\, q(t) = q\right]$

Step 1: Drift of exponential test function

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

Arrival Departure

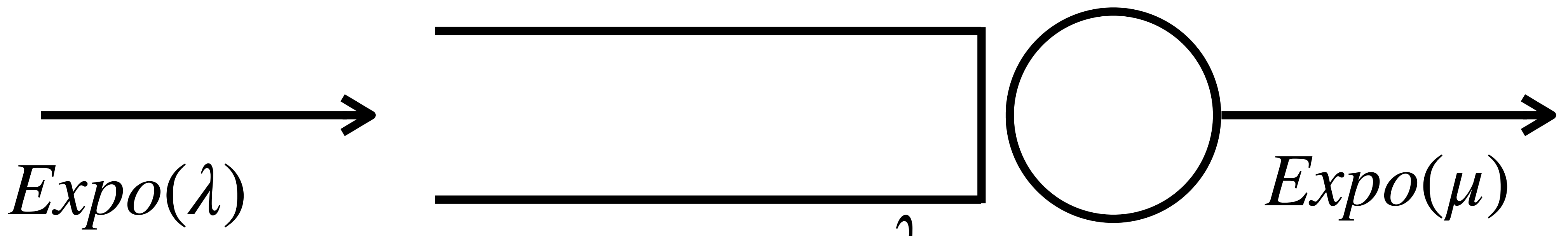
$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$

$$= e^{\theta q}\left(e^{-\theta} - 1\right)\left(\mu - \lambda e^{\theta}\right) - \mu\left(e^{-\theta} - 1\right) \underbrace{e^{\theta q} 1_{\{q=0\}}}_{= 1_{\{q=0\}}}$$

Step 2: Set drift to zero

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

Transform Method for $M/M/1$ queue



Load: $\rho = \dfrac{\lambda}{\mu}$

Drift:
$\Delta\varphi_\theta(q) = \mathbb{E}\left[e^{\theta q(t+)} - e^{\theta q} \middle| \, q(t) = q\right]$

Step 1: Drift of exponential test function

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

Arrival Departure

$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$

$$= e^{\theta q}\left(e^{-\theta} - 1\right)\left(\mu - \lambda e^{\theta}\right) - \mu\left(e^{-\theta} - 1\right) \underbrace{e^{\theta q} 1_{\{q=0\}}}_{= 1_{\{q=0\}}}$$

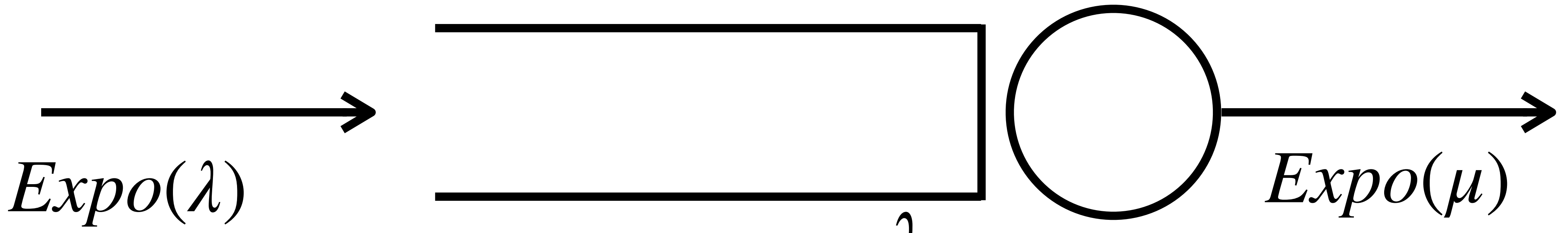
Step 2: Set drift to zero

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

Step 3: Set $\theta = 0$

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

Transform Method for $M/M/1$ queue



Load: $\rho = \frac{\lambda}{\mu}$

Drift:
$\Delta\varphi_\theta(q) = \mathbb{E}\left[e^{\theta q(t+)} - e^{\theta q} \,\middle|\, q(t) = q\right]$

Step 1: Drift of exponential test function

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

Arrival — Departure

$$\Delta\varphi_\theta(q) = \lambda(e^{\theta(q+1)} - e^{\theta q}) + \mu 1_{\{q>0\}}(e^{\theta(q-1)} - e^{\theta q})$$

$$= e^{\theta q}\left(e^{-\theta} - 1\right)\left(\mu - \lambda e^{\theta}\right) - \mu\left(e^{-\theta} - 1\right) e^{\theta q} 1_{\{q=0\}}$$

$$e^{\theta q} 1_{\{q=0\}} = 1_{\{q=0\}}$$

Step 2: Set drift to zero

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

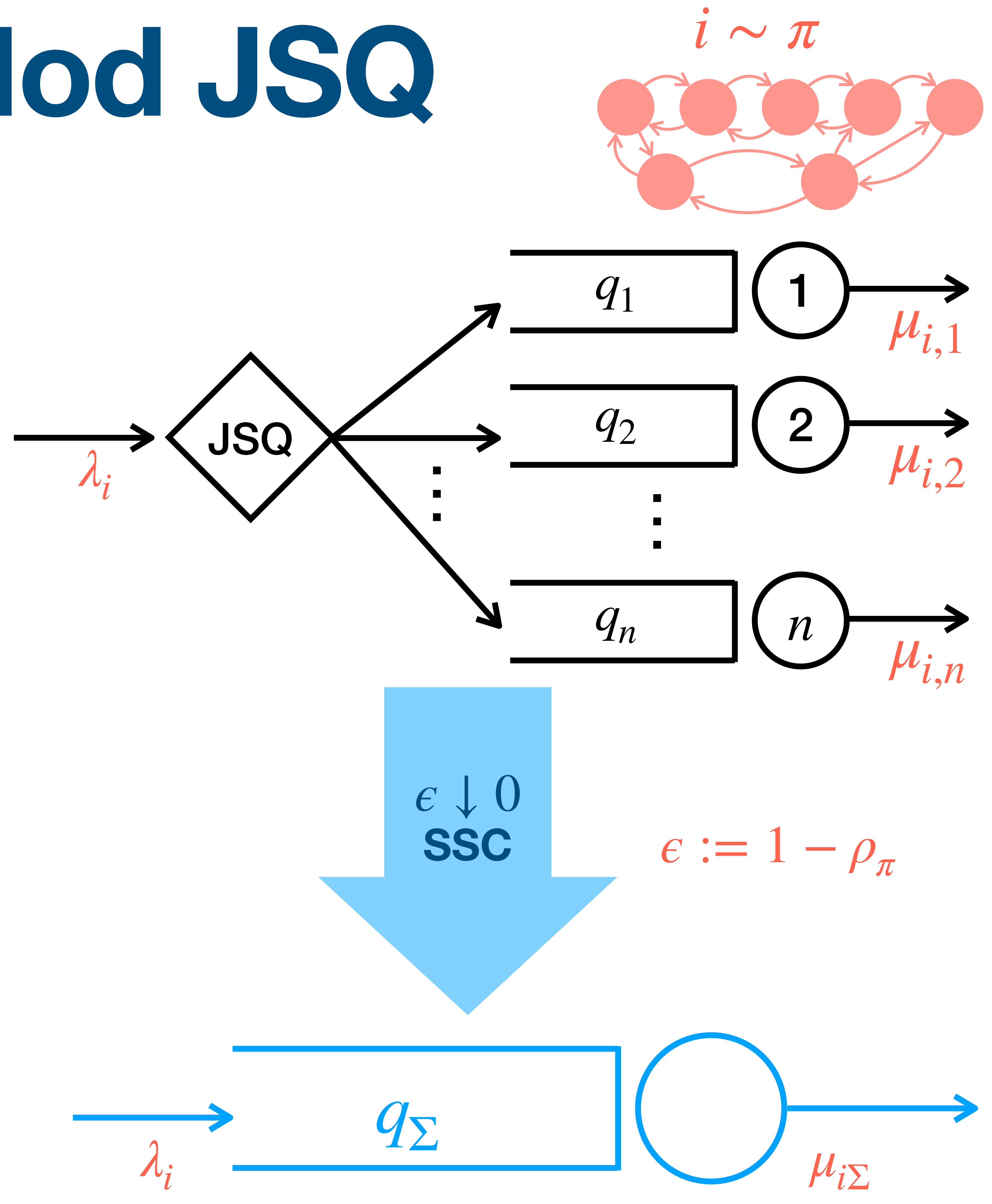
Step 3: Set $\theta = 0$

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

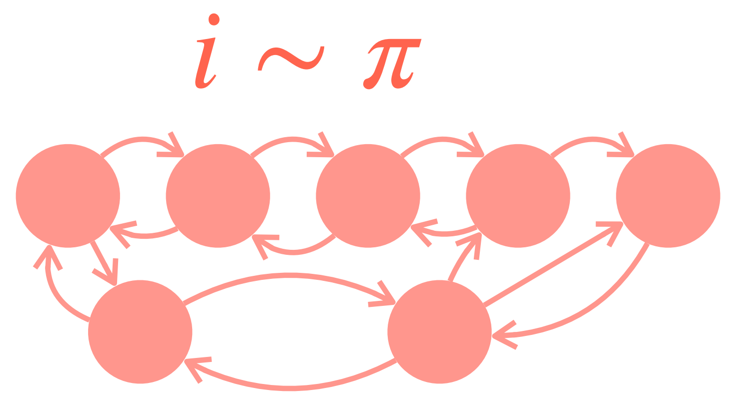
$$\mathbb{E}\left[e^{\theta q}\right] = \frac{1-\rho}{1-\rho e^{\theta}}$$

$\Longleftrightarrow$ $q \sim$ Geometric$(1-\rho)$

Transform Method Markov-Mod JSQ

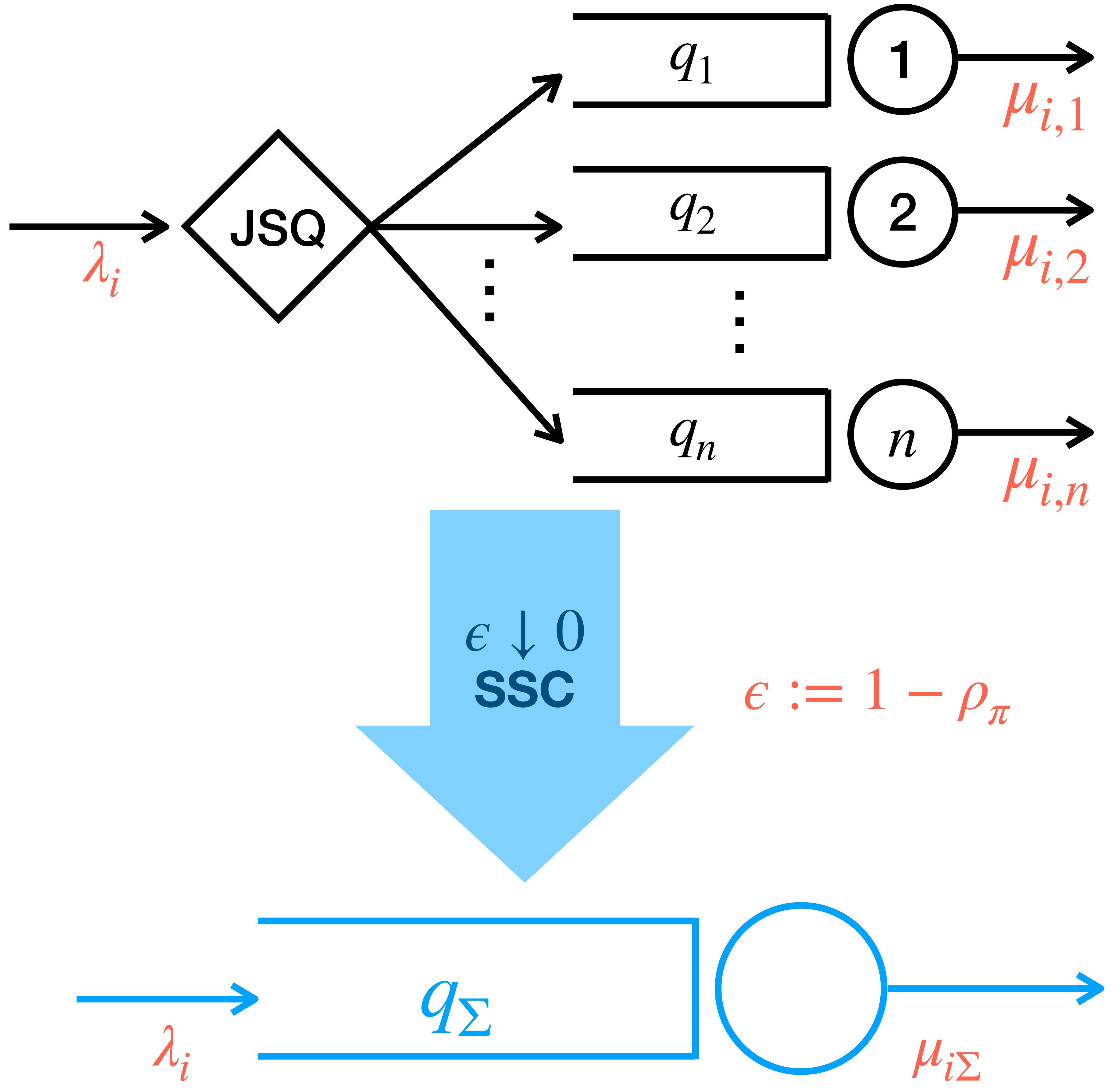


Transform Method Markov-Mod JSQ

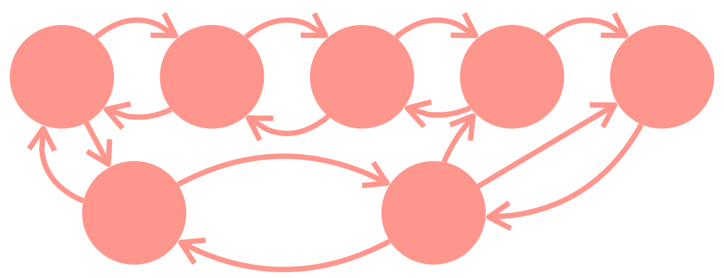


Step 1: Drift of exponential test function

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



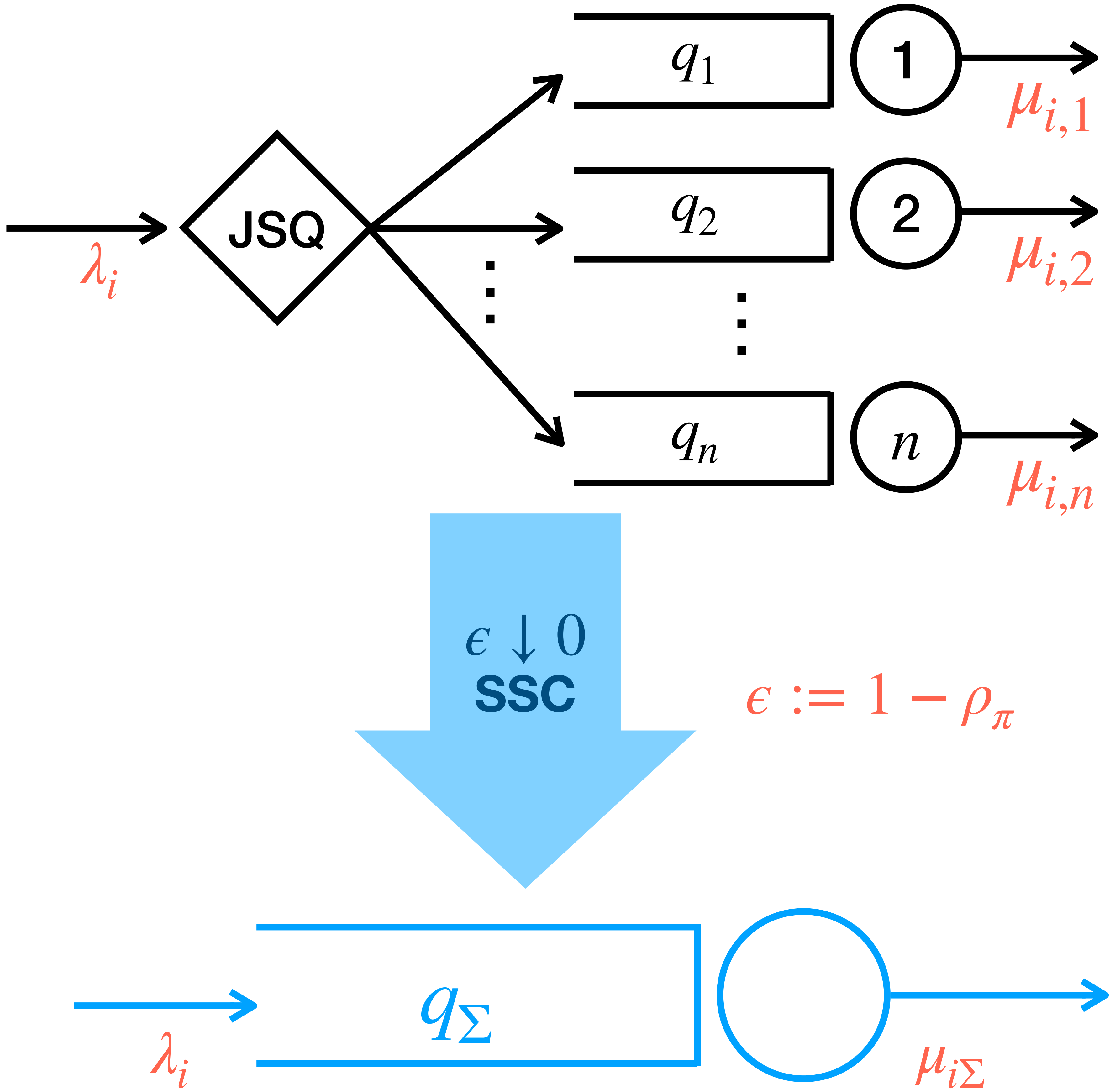
Transform Method Markov-Mod JSQ



Step 1: Drift of exponential test function

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

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



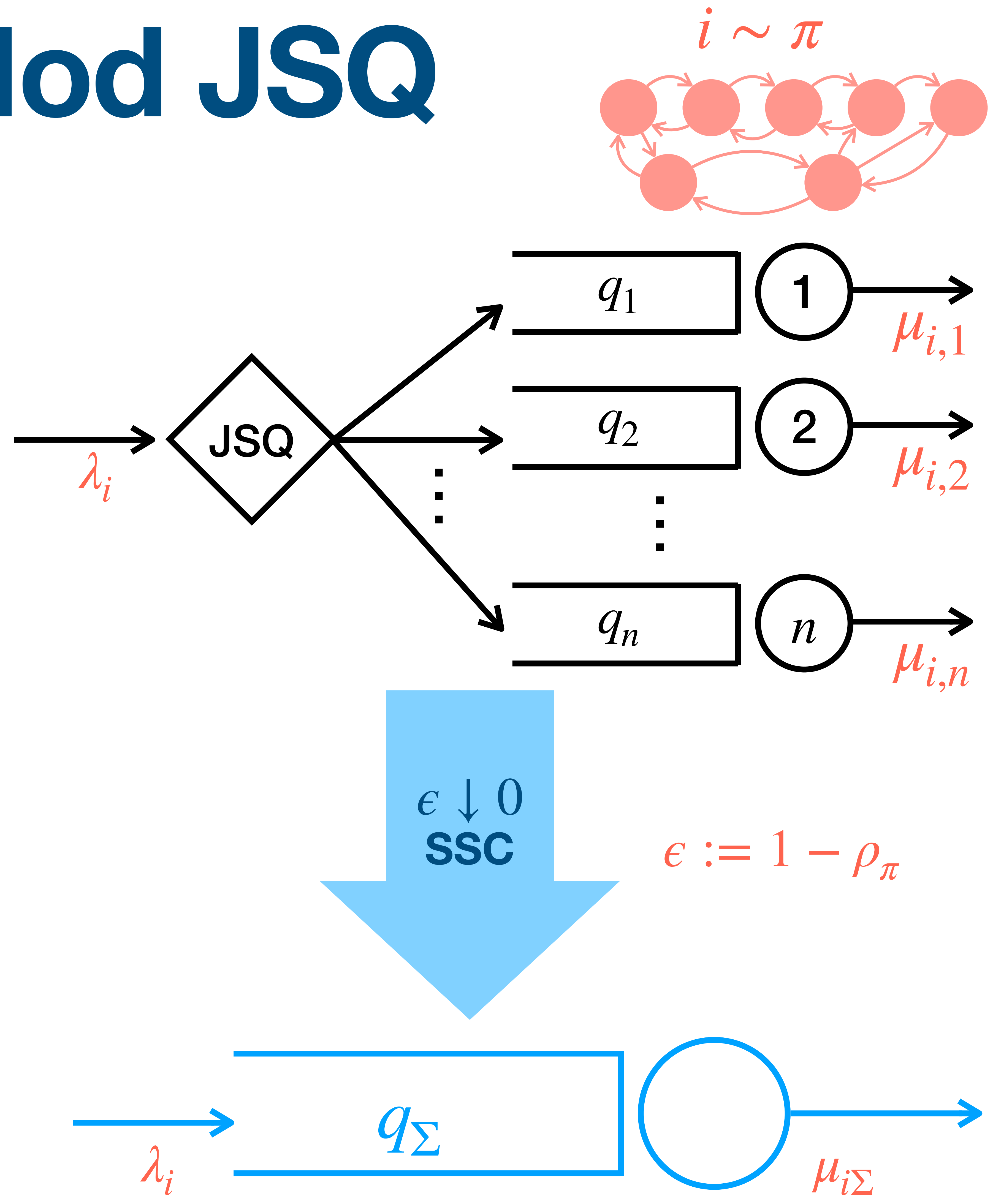
Transform Method Markov-Mod JSQ

Step 1: Drift of exponential test function

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

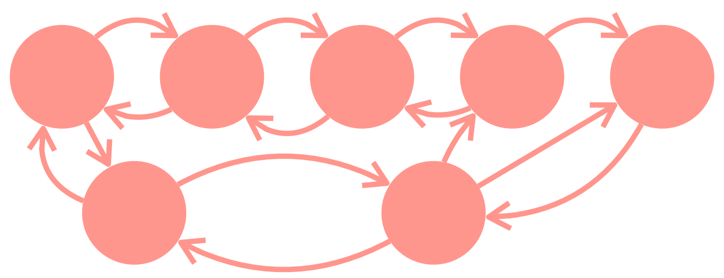
$$\Delta \varphi_s(i, \boldsymbol{q}) = \left(e^{-s\epsilon} - 1\right) e^{-s\epsilon q_\Sigma} \left(\lambda_i - \mu_{i\Sigma} e^{s\epsilon}\right)$$

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



Transform Method Markov-Mod JSQ

$i \sim \pi$



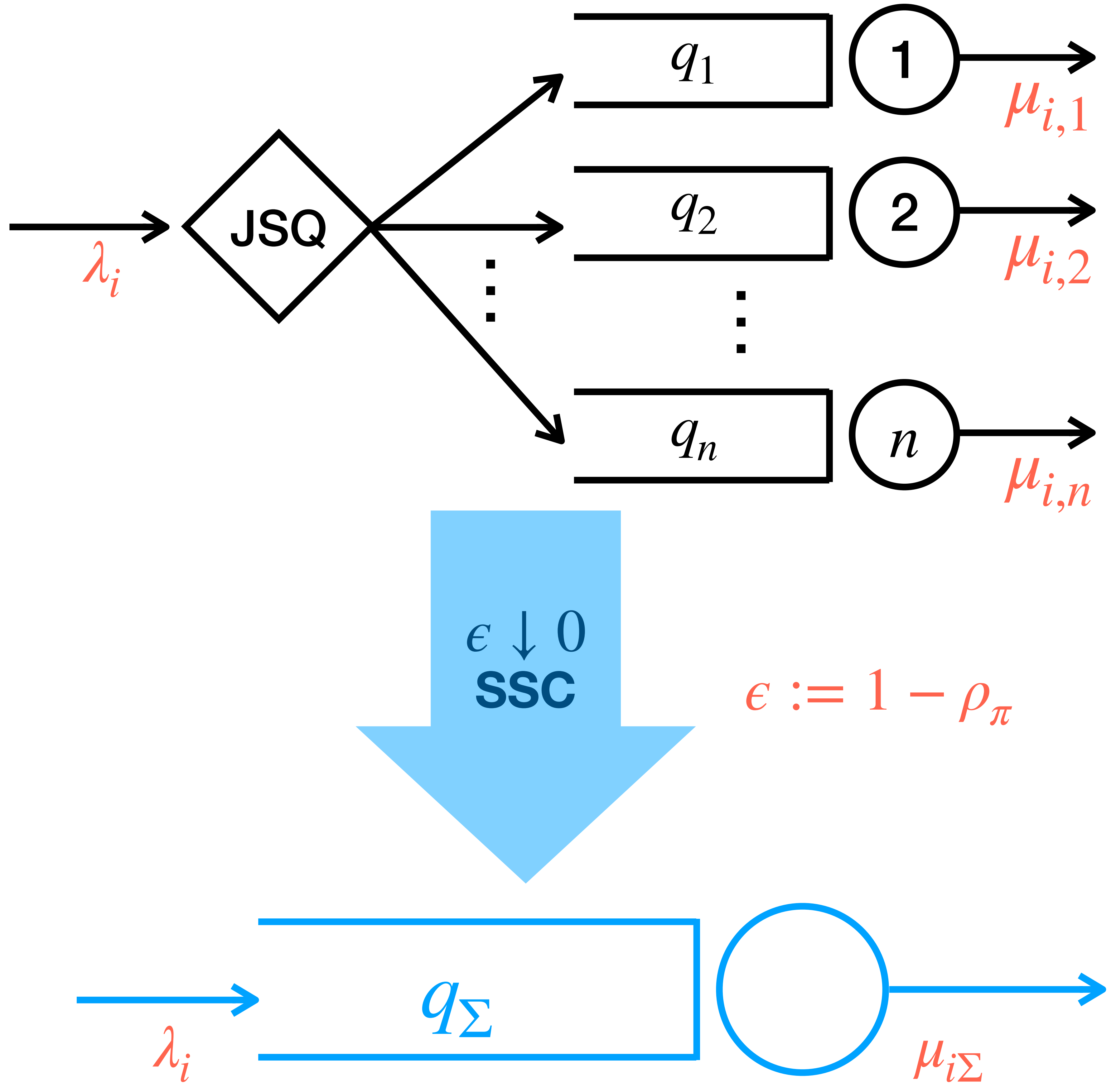
Step 1: Drift of exponential test function

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

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

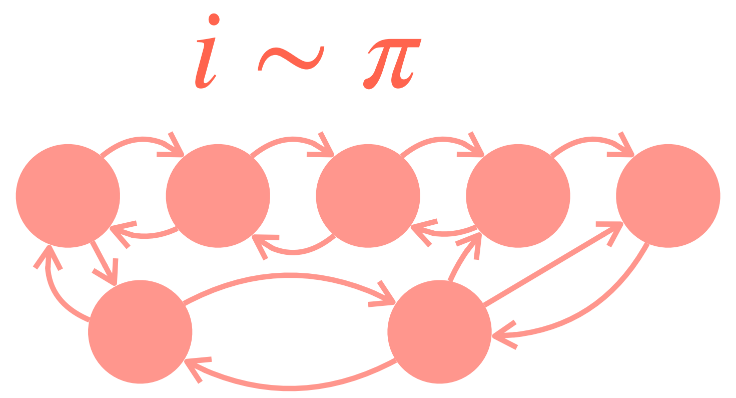
$$-\left(e^{-s\epsilon} - 1\right) e^{-s\epsilon q_\Sigma}\Big(\sum_j \mu_{ij} 1_{\{q_j=0\}}\Big)$$

$$\approx \sum_j \mu_{ij}\, e^{-s\epsilon n q_j} 1_{\{q_j=0\}} \quad \text{(SSC)}$$



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

Transform Method Markov-Mod JSQ



Step 1: Drift of exponential test function

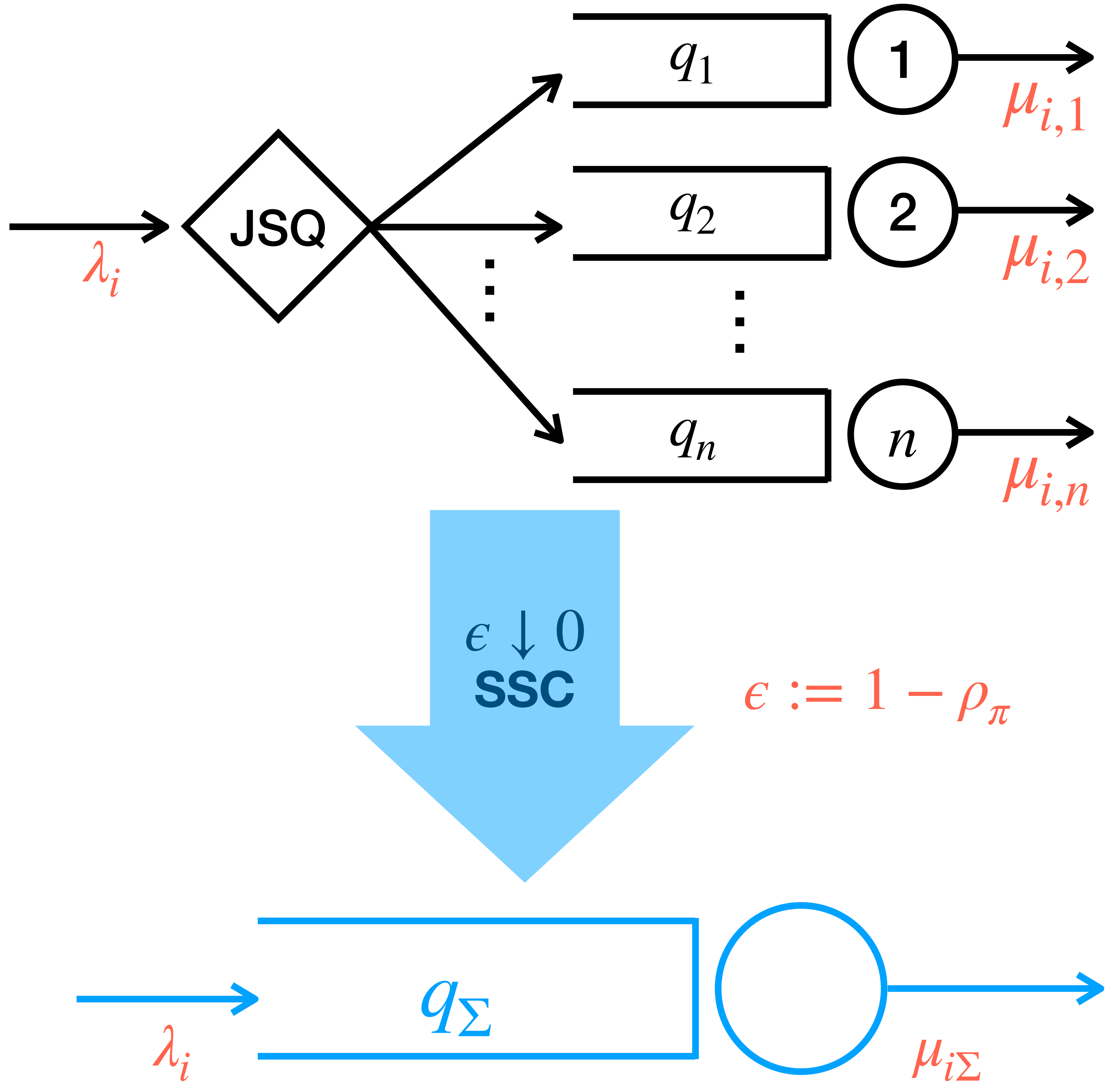
$$\varphi_s(i, \boldsymbol{q}) = e^{-s\epsilon q_\Sigma}, \; s > 0$$

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

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

$$\approx \sum_j \mu_{ij} e^{-s\epsilon n q_j} 1_{\{q_j=0\}} \qquad \text{(SSC)}$$

$$= \sum_j \mu_{ij} 1_{\{q_j=0\}}$$



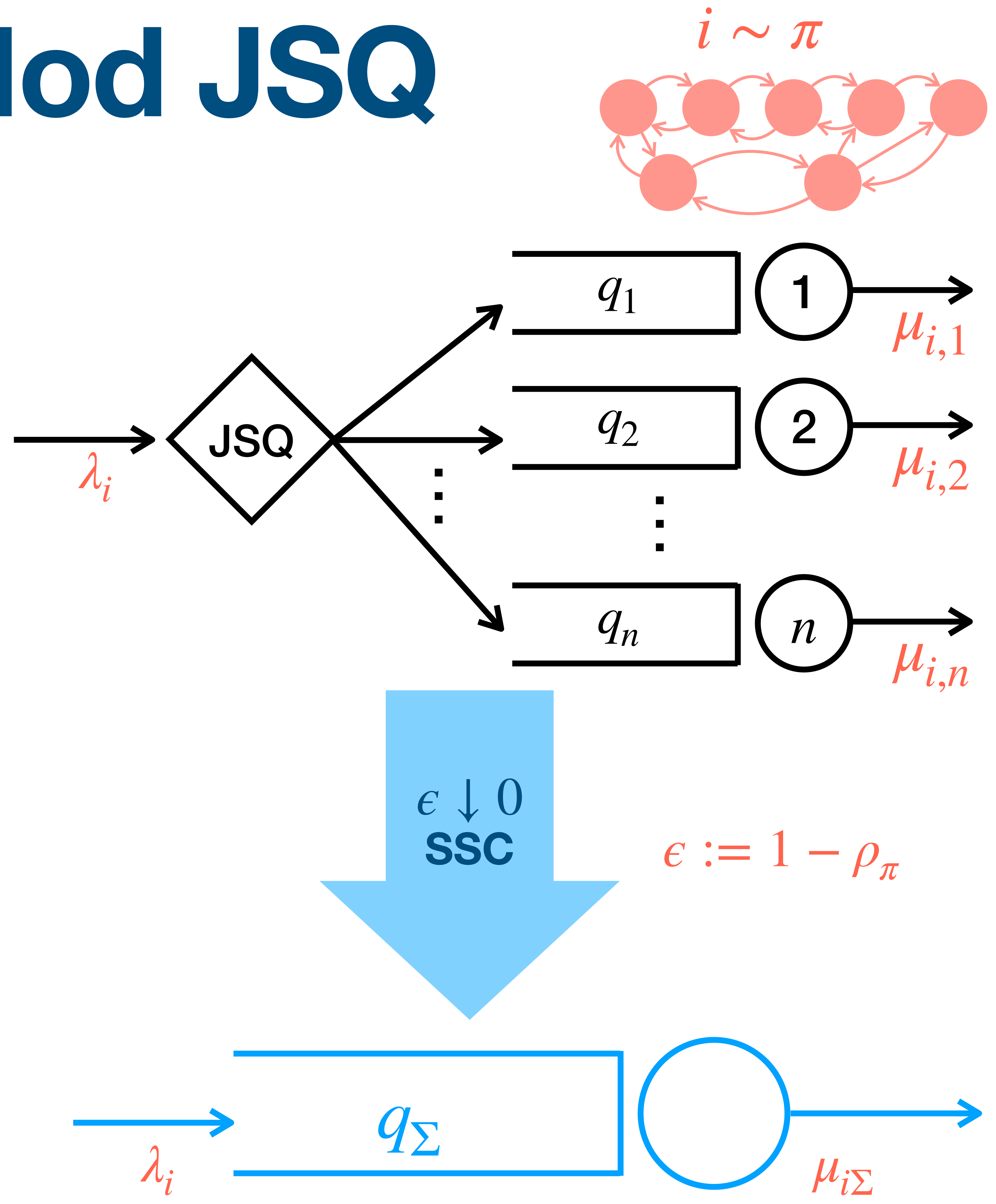
Transform Method Markov-Mod JSQ

Step 1: Drift of exponential test function

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

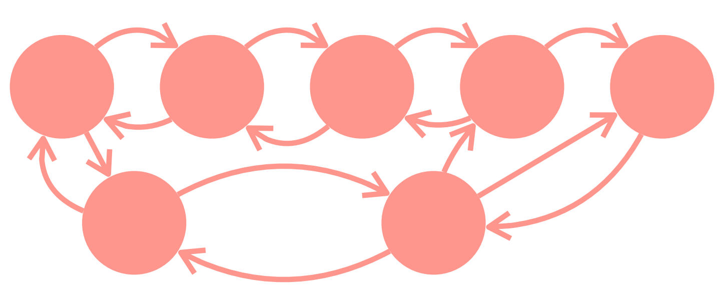
$$\Delta\varphi_s(i, \boldsymbol{q}) = \left(e^{-s\epsilon} - 1\right) e^{-s\epsilon q_\Sigma}\left(\lambda_i - \mu_{i\Sigma}\, e^{s\epsilon}\right)$$

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



Transform Method Markov-Mod JSQ

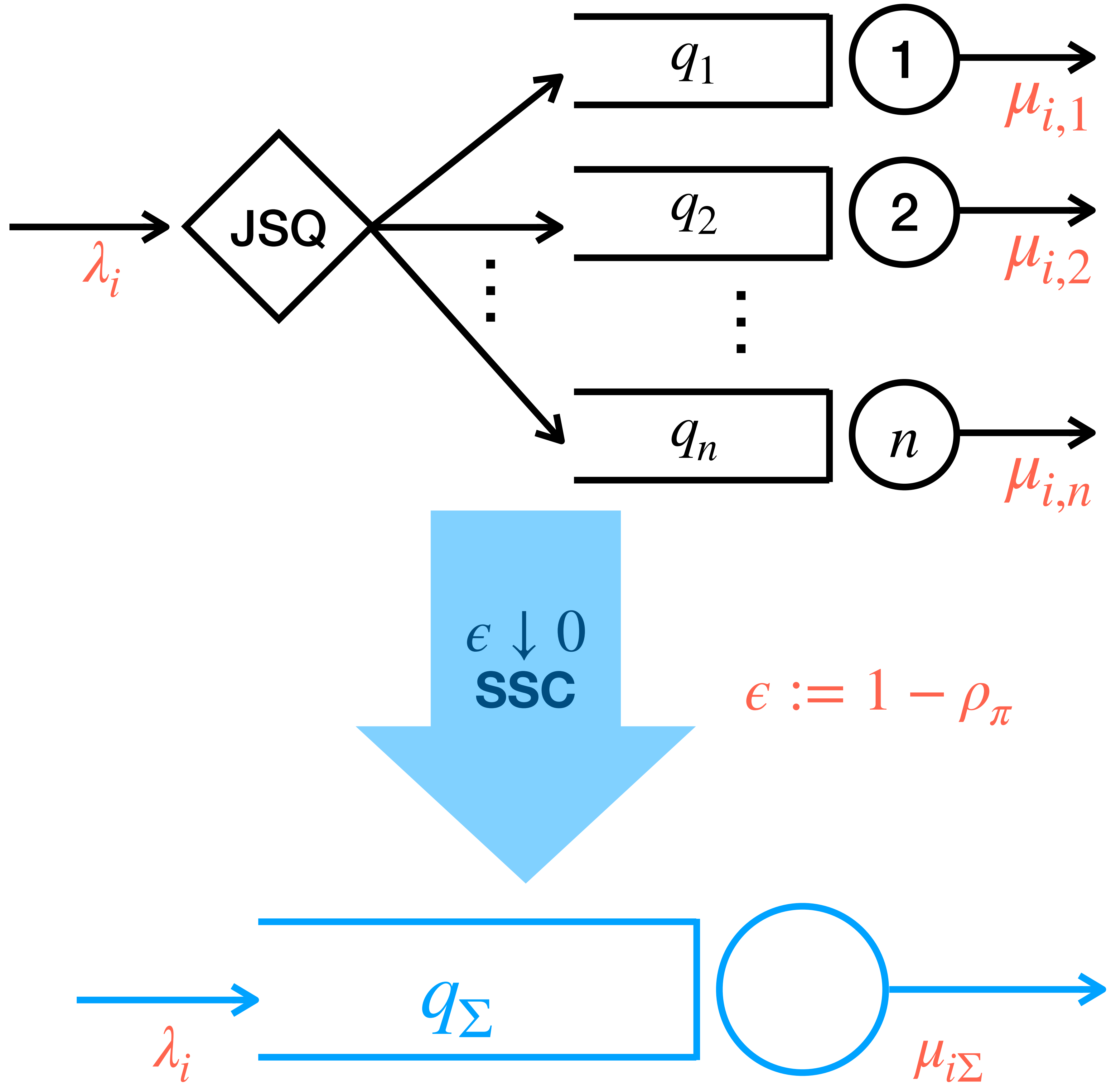
$i \sim \pi$



Step 1: Drift of exponential test function

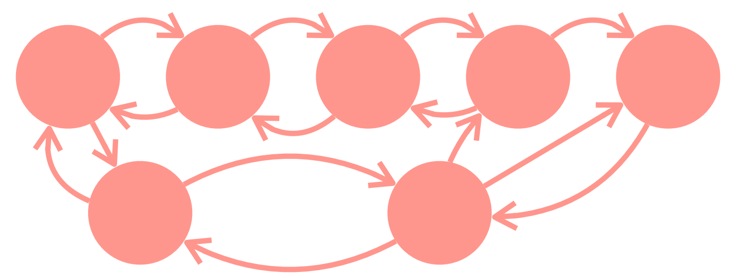
$\varphi_s(i, \boldsymbol{q}) = e^{-s\epsilon q_\Sigma}$, $s > 0$

$$\Delta\varphi_s(i, \boldsymbol{q}) = \left(e^{-s\epsilon} - 1\right) e^{-s\epsilon q_\Sigma}\left(\lambda_i - \mu_{i\Sigma}\, e^{s\epsilon}\right) - \left(1 - e^{s\epsilon}\right)\underbrace{\left(\sum_j \mu_{ij} 1_{\{q_j=0\}}\right)}_{\mathbb{E}\left[\sum_j \mu_{ij} 1_{\{q_j=0\}}\right] = \mu_\Sigma\, \epsilon}$$



Transform Method Markov-Mod JSQ

$i \sim \pi$



Step 1: Drift of exponential test function

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

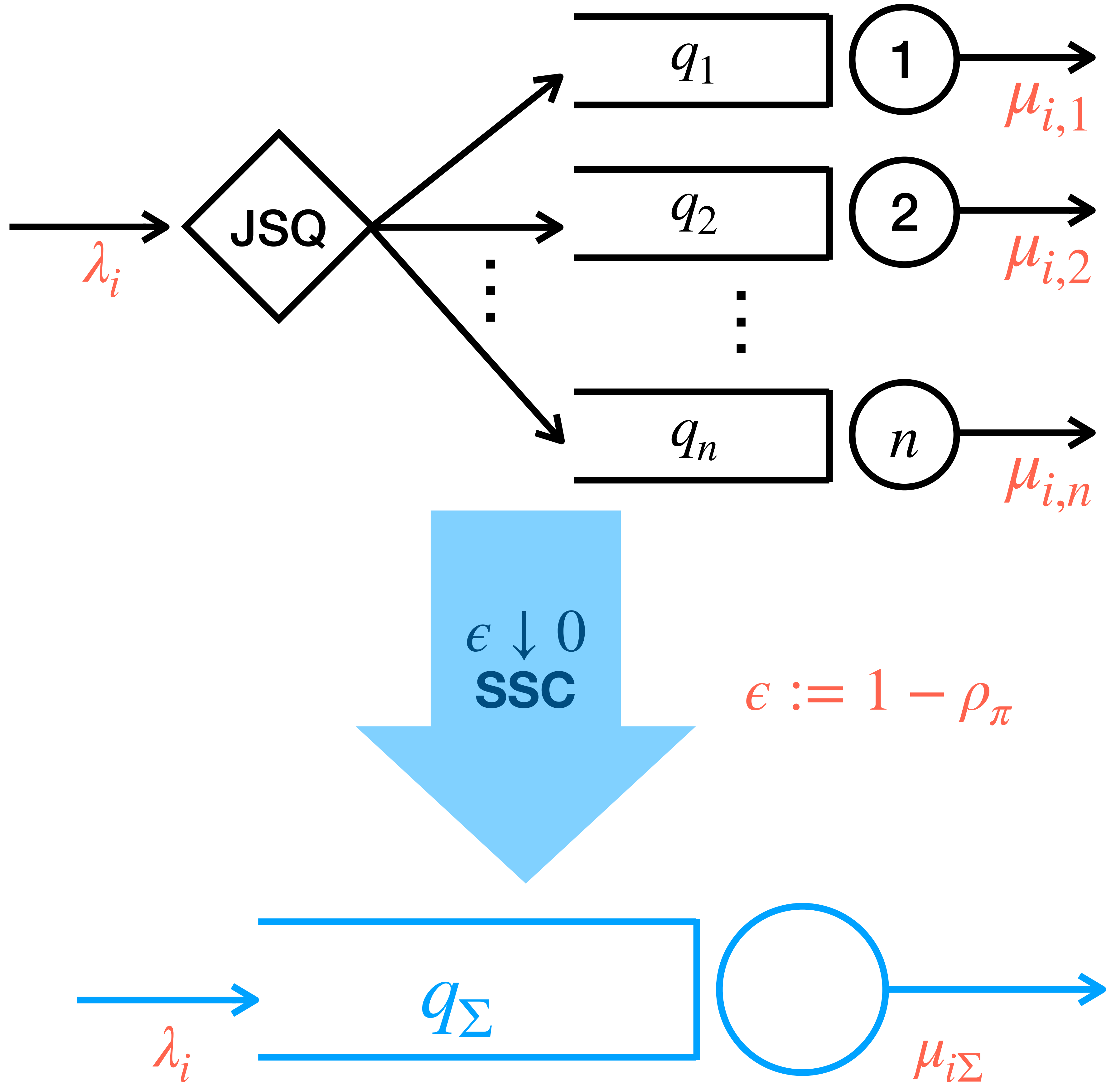
Not independent!

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

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

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

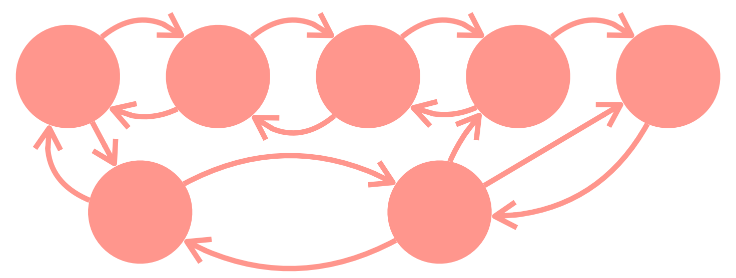
$$\mathbb{E}\Big[\sum_j \mu_{ij} 1_{\{q_j=0\}}\Big] = \mu_\Sigma\, \epsilon$$



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

Transform Method Markov-Mod JSQ

$i \sim \pi$



Step 1: Drift of exponential test function

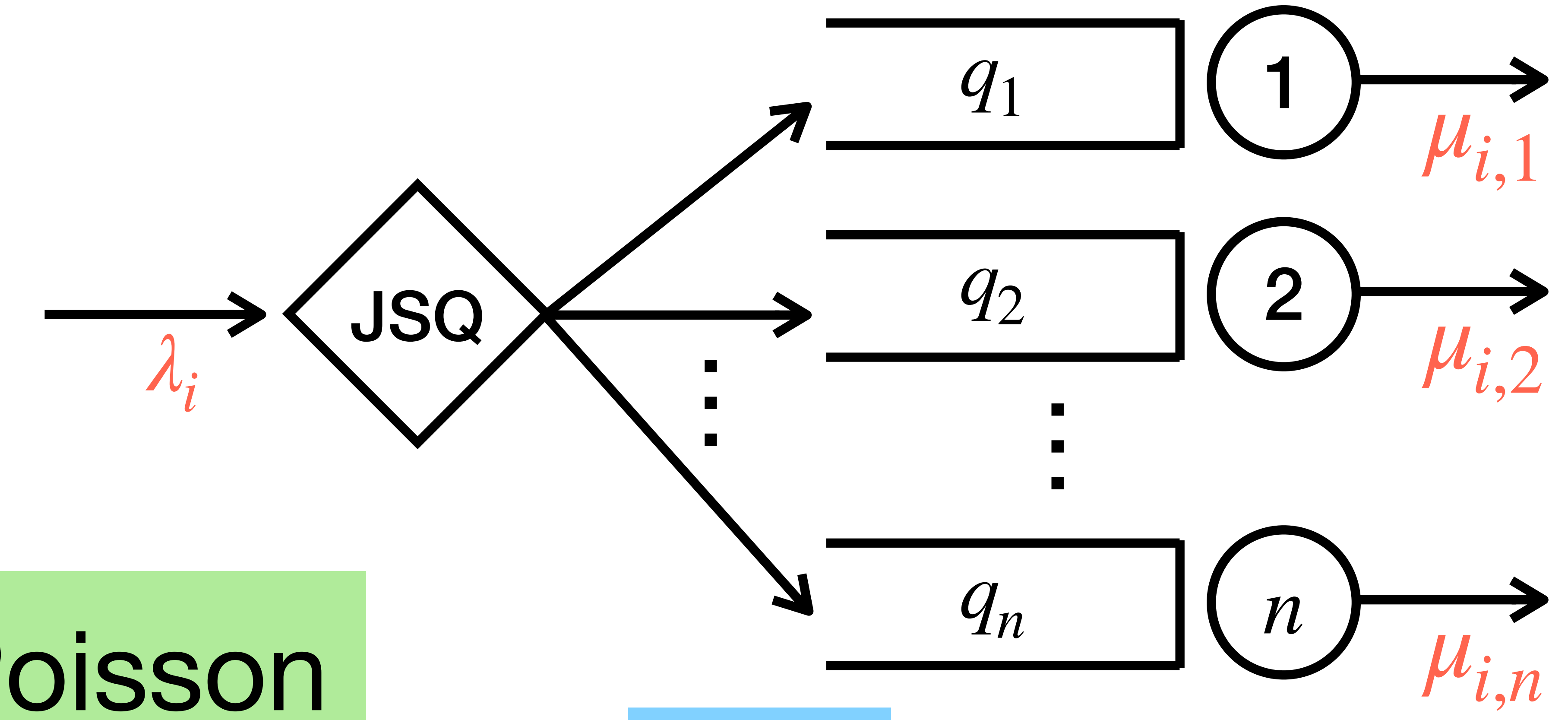
$$\varphi_s(i,\boldsymbol{q}) = e^{-s\epsilon q_\Sigma},\ s > 0$$

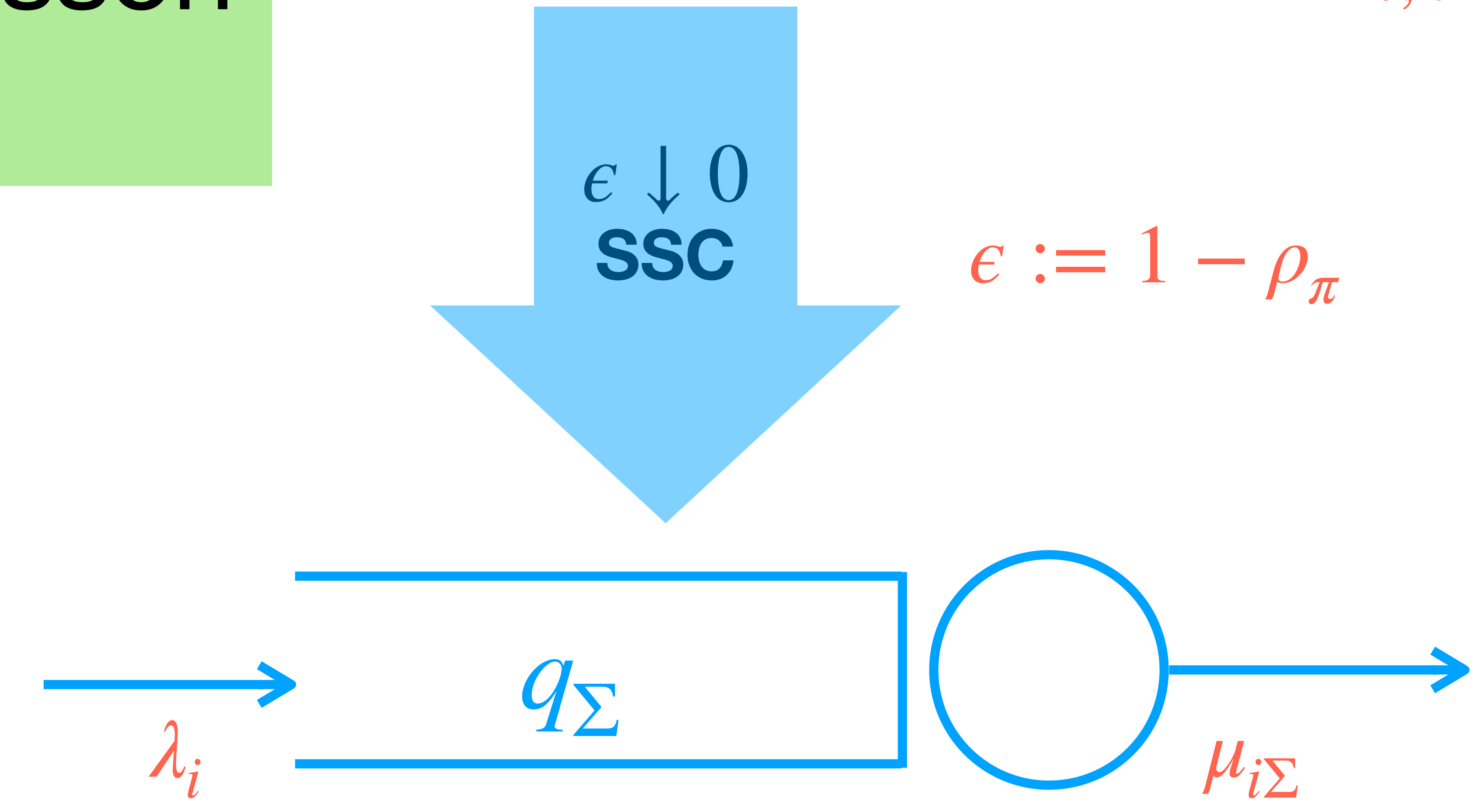
Not independent!

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

$$\Delta\varphi_s(i,\boldsymbol{q}) = \left(e^{-s\epsilon} - 1\right) e^{-s\epsilon q_\Sigma}\left(\lambda_i - \mu_{i\Sigma}\, e^{s\epsilon}\right) - \left(1 - e^{s\epsilon}\right)\Big(\underbrace{\sum_j \mu_{ij} 1_{\{q_j=0\}}}_{\mathbb{E}\left[\sum_j \mu_{ij} 1_{\{q_j=0\}}\right] = \mu_\Sigma \epsilon}\Big)$$

We use the Poisson equation!

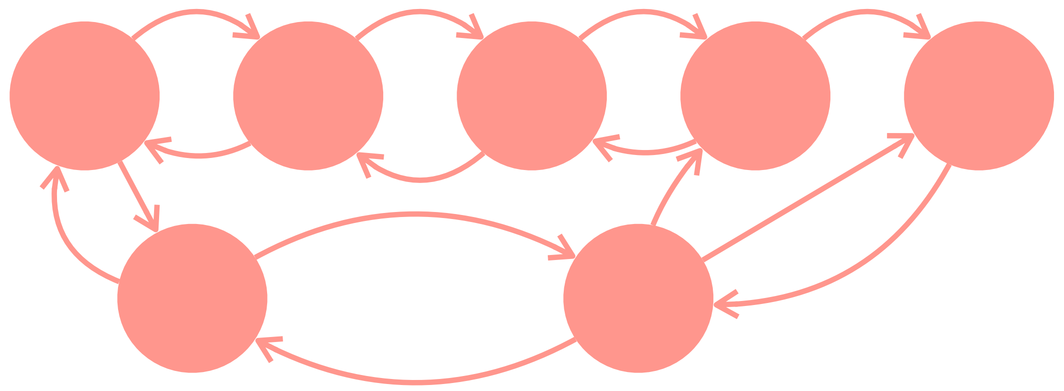




$\epsilon := 1 - \rho_\pi$

Poisson Equation for Transform Method

$Z(t) \sim$

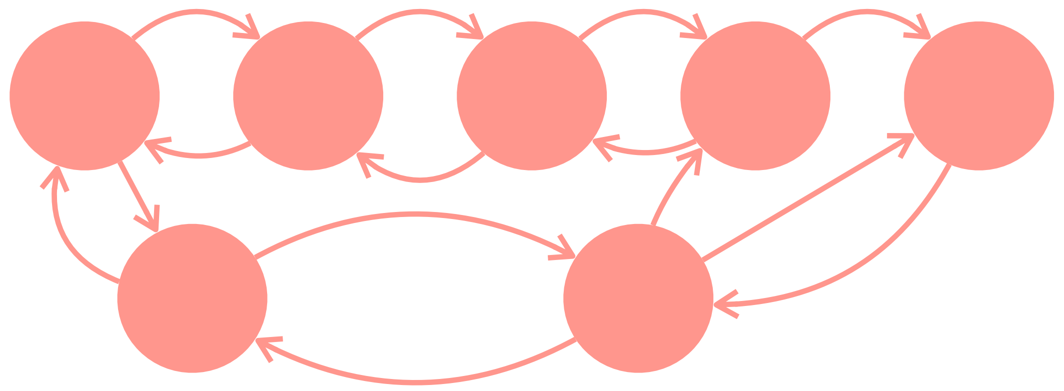


Poisson Equation for Transform Method

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

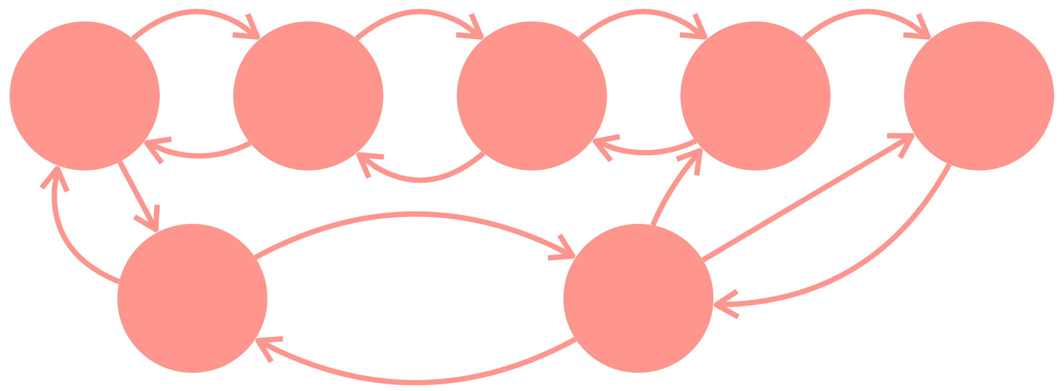
$$V_f(i) = \frac{f(i) - \bar{f}}{\alpha_{i\cdot}} + \sum_{i' \neq i} \frac{\alpha_{ii'}}{\alpha_{i\cdot}} V_f(i')$$

$Z(t) \sim$



Poisson Equation for Transform Method

$Z(t) \sim$



Poisson equation:

Let $\{Z(t)\}_t$ be a CTMC with countable state space $\mathcal{Z}$ and transition rates $\alpha_{i,i'}$

Consider a function $f : \mathcal{Z} \to \mathbb{R}$ and let $\bar{f} = \mathbb{E}[f(Z)]$. Then, there exists a function $V_f : \mathcal{Z} \to \mathbb{R}$ such that

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

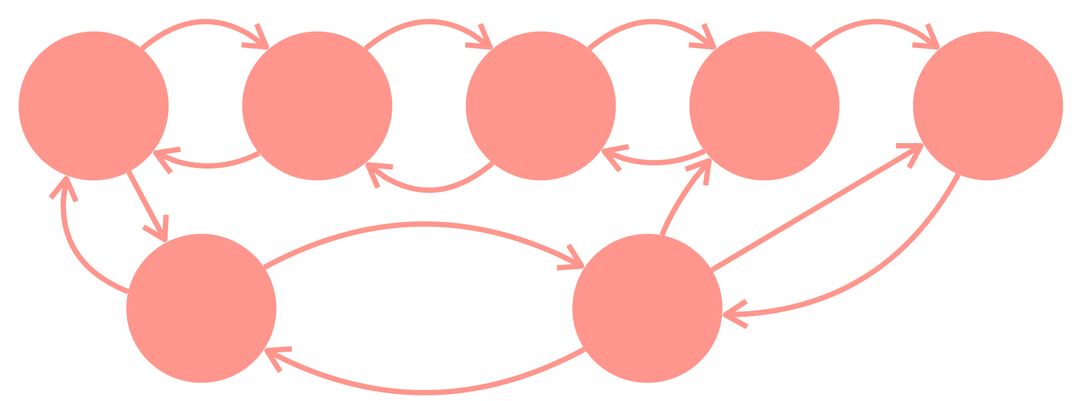
Theorem [HL, Grosof '25]:

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

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

Poisson Equation for Transform Method

$Z(t) \sim$



Poisson equation:

Let $\{Z(t)\}_t$ be a CTMC with countable state space $\mathscr{Z}$ and transition rates $\alpha_{i,i'}$

Consider a function $f : \mathscr{Z} \to \mathbb{R}$ and let $\bar{f} = \mathbb{E}[f(Z)]$. Then, there exists a function $V_f : \mathscr{Z} \to \mathbb{R}$ such that

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

Theorem [HL, Grosof '25]:

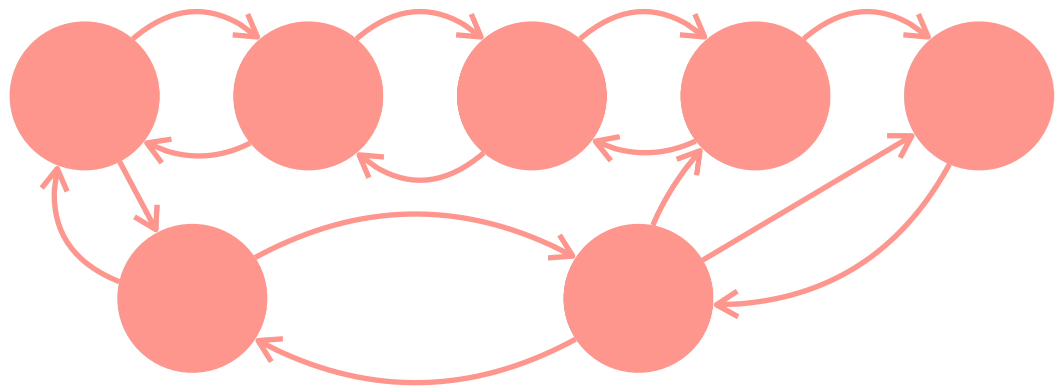
For any function $f : \mathscr{Z} \to \mathbb{R}$ such that $V_f(i)$ exists and $\mathbb{E}\left[|V_f(i)|^{1+\eta}\right] \leq C$ for some $\eta > 0$,

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

✓Separate expectation of product

Poisson Equation for Transform Method

$Z(t) \sim$



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

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

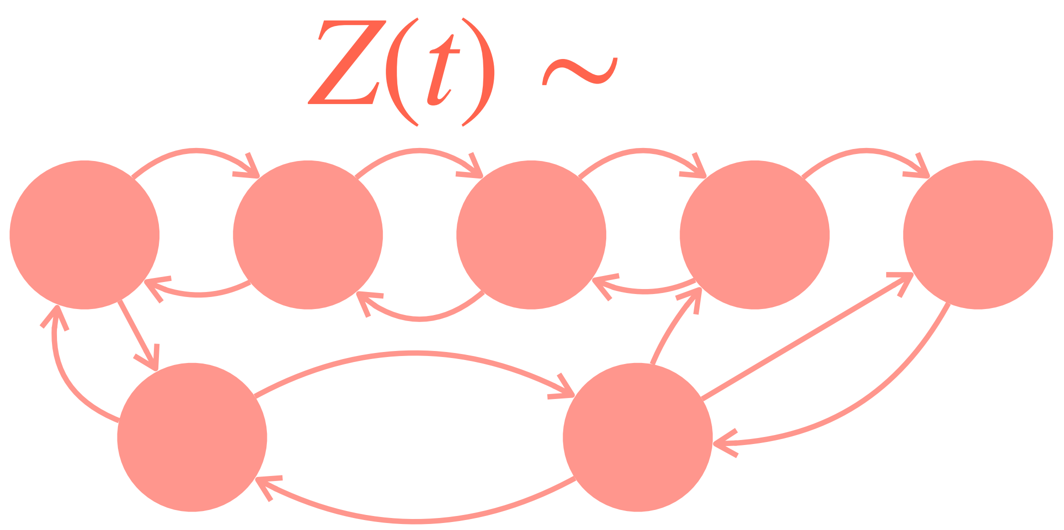
Theorem [HL, Grosof '25]:
For any function $f : \mathscr{Z} \to \mathbb{R}$ such that $V_f(i)$ exists and $\mathbb{E}\left[|V_f(i)|^{1+\eta}\right] \leq C$ for some $\eta > 0$,

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

✓Separate expectation of product

Solution to
Poisson equation

Poisson Equation for Transform Method



Poisson equation:

Let $\{Z(t)\}_t$ be a CTMC with countable state space $\mathscr{Z}$ and transition rates $\alpha_{i,i'}$

Consider a function $f : \mathscr{Z} \to \mathbb{R}$ and let $\bar{f} = \mathbb{E}[f(Z)]$. Then, there exists a function $V_f : \mathscr{Z} \to \mathbb{R}$ such that

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

Theorem [HL, Grosof '25]:

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

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

✓ Separate expectation of product

Solution to Poisson equation

Error term

Applying the Poisson Equation

Applying the Poisson Equation

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

Applying the Poisson Equation

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

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

Applying the Poisson Equation

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

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

Applying the Poisson Equation

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

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

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

Applying the Poisson Equation

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

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

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

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

Applying the Poisson Equation

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

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

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

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

Applying the Poisson Equation

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

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

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

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

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

Applying the Poisson Equation

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

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

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

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

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

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

Applying the Poisson Equation

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

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

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

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

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

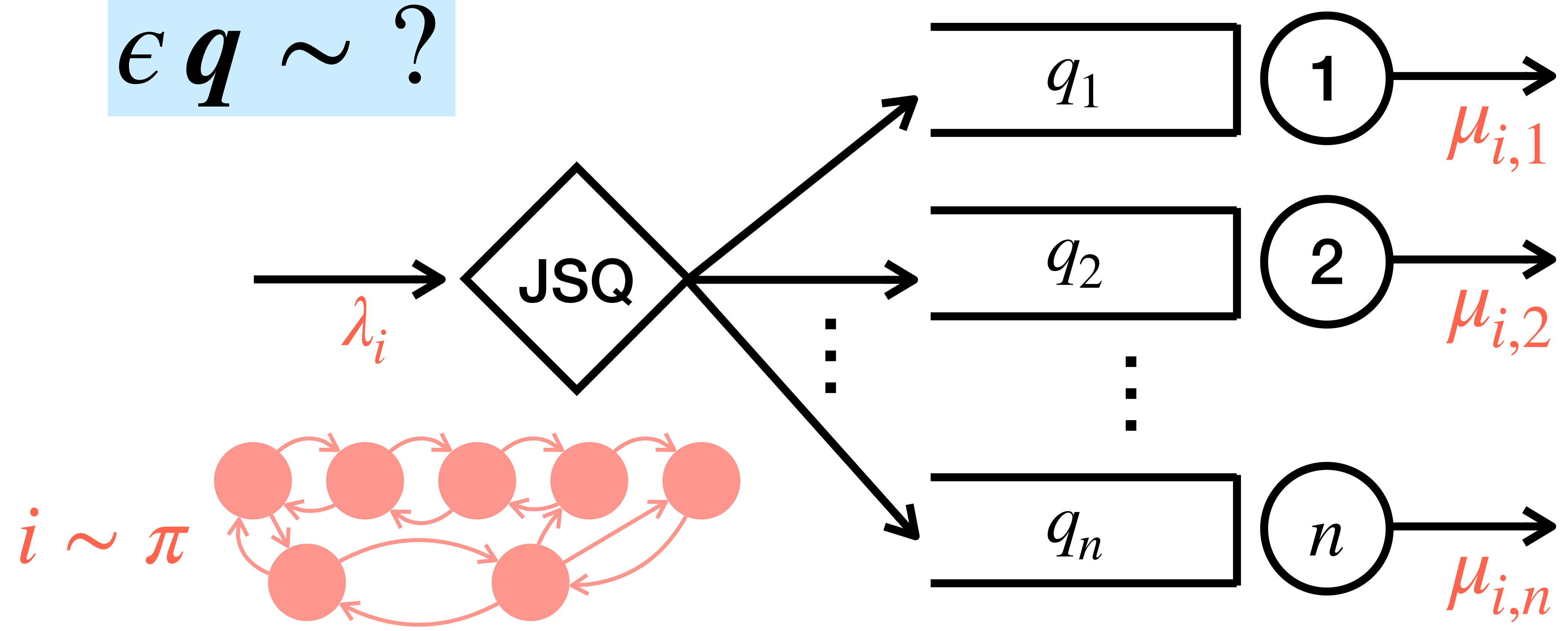
$$\mathbb{E}[e^{-s\epsilon\, q_\Sigma}V_h(i)(\mu_{i\Sigma} - \lambda_i)] = \mathbb{E}[e^{-s\epsilon q_\Sigma}]\,\mathbb{E}[k(i)] + O(\epsilon)$$

Then, put everything together. ■

Key Takeaways

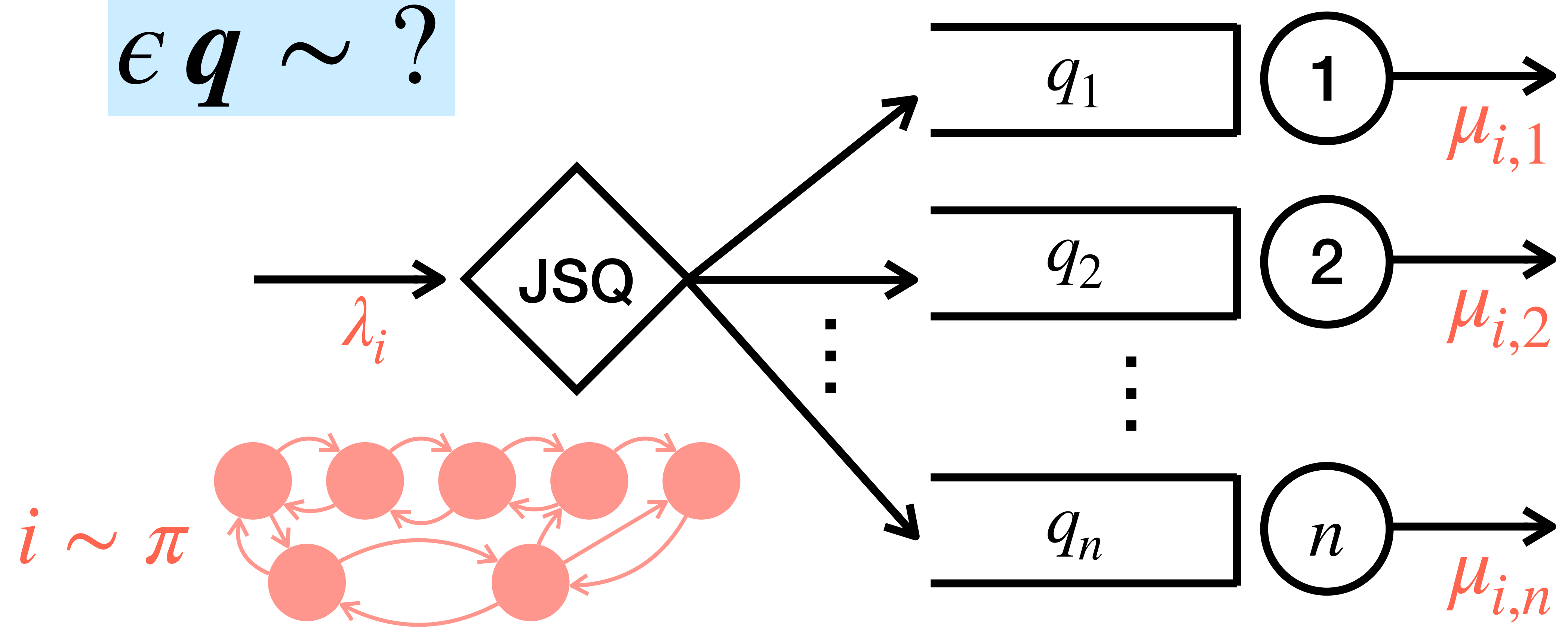
Key Takeaways

$\epsilon \boldsymbol{q} \sim ?$



Key Takeaways

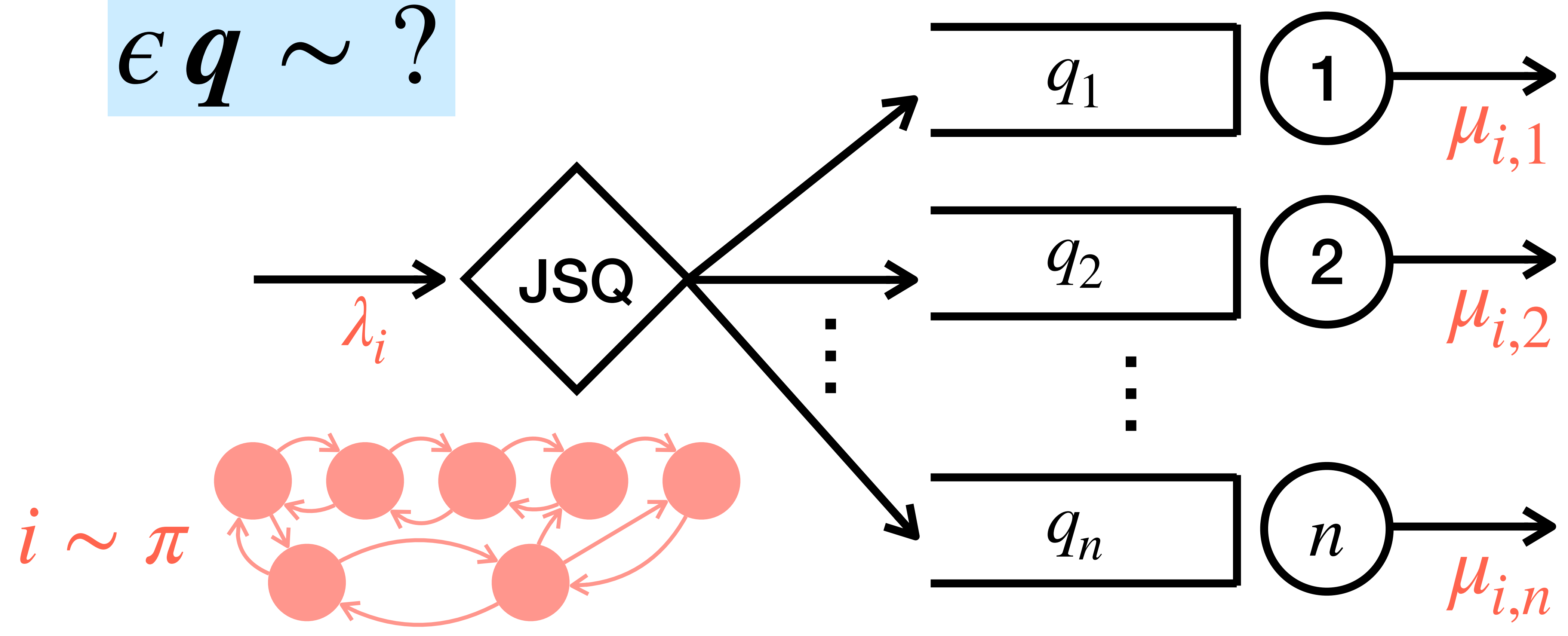
$\epsilon \boldsymbol{q} \sim ?$



Step 1: State Space Collapse

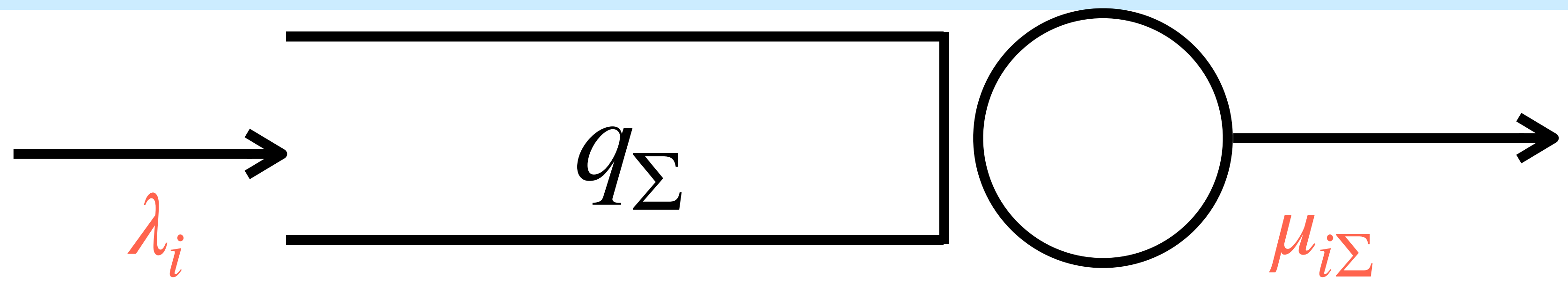
Key Takeaways

$\epsilon \boldsymbol{q} \sim ?$

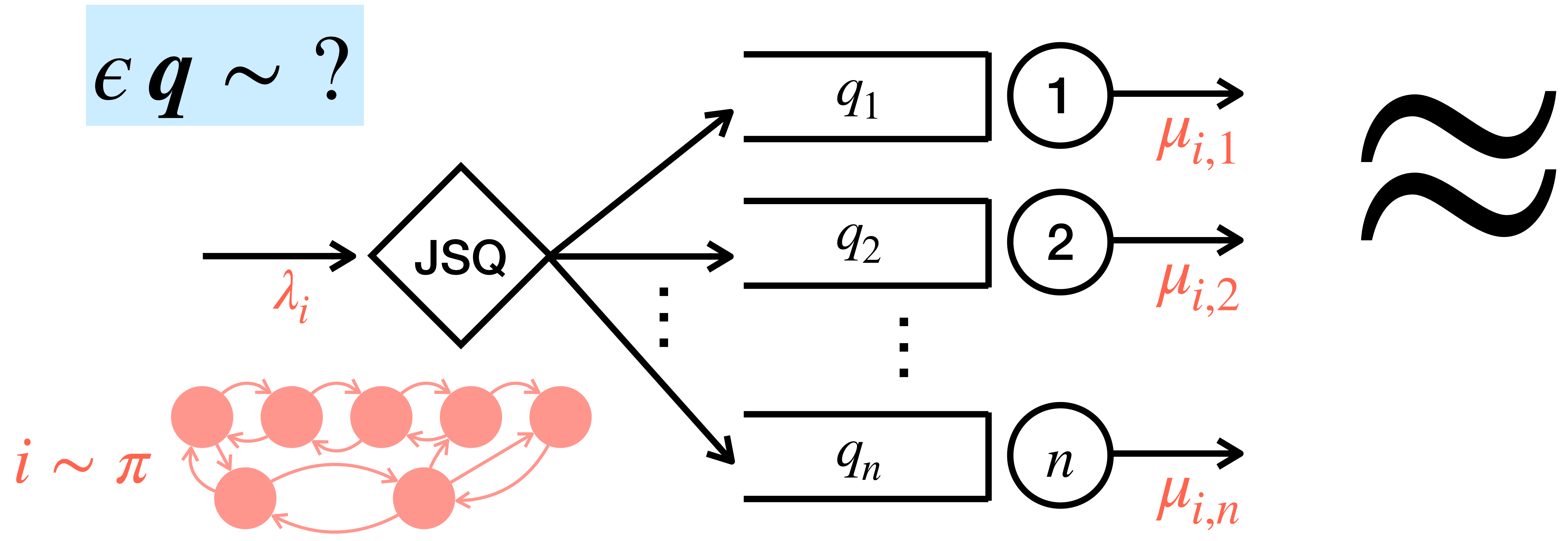


$\approx$

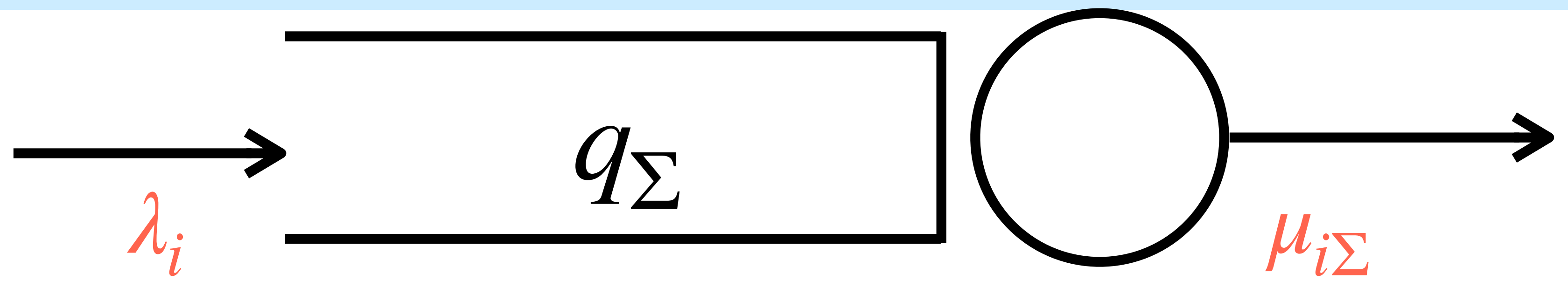
Step 1: State Space Collapse



Key Takeaways



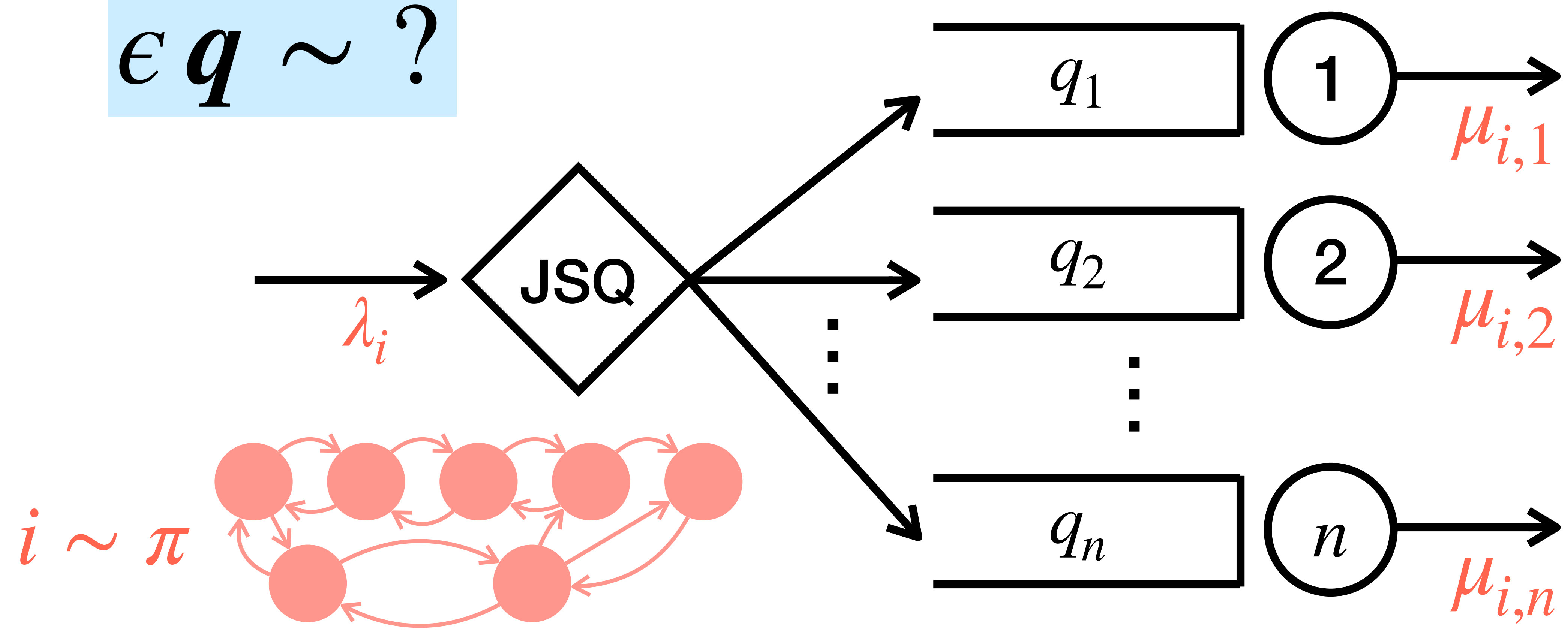
Step 1: State Space Collapse



✓ Only need λ_i large enough so JSQ balances queue lengths for all i

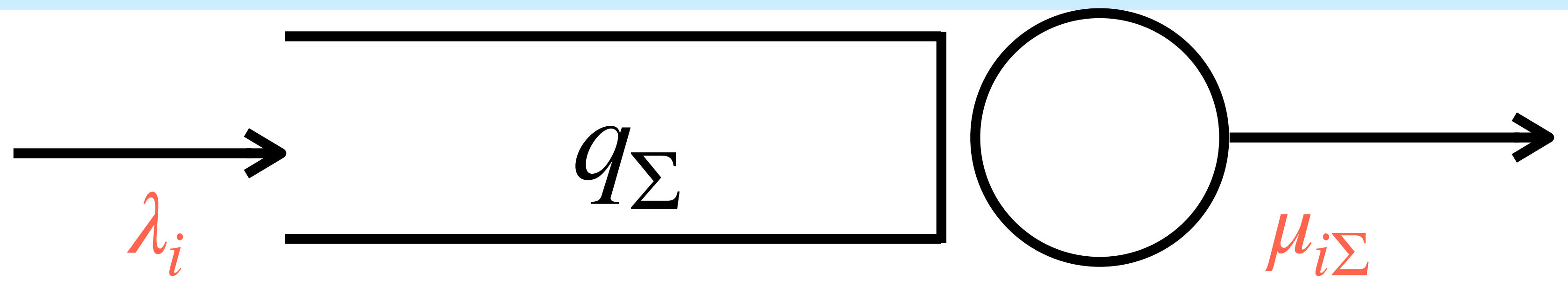
Key Takeaways

$\epsilon \boldsymbol{q} \sim ?$



$\approx$

Step 1: State Space Collapse

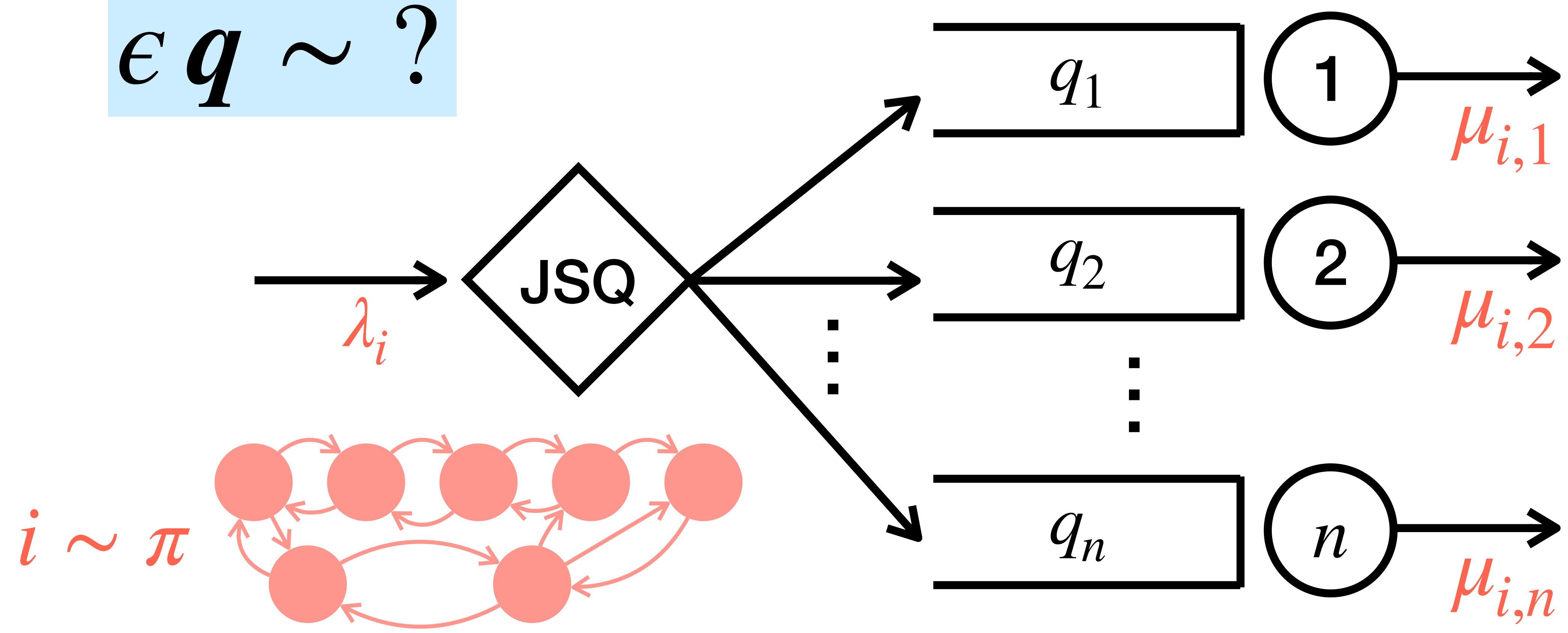


✓ Only need λ_i large enough so JSQ balances queue lengths for all i

✓ Queue can be unstable for some i

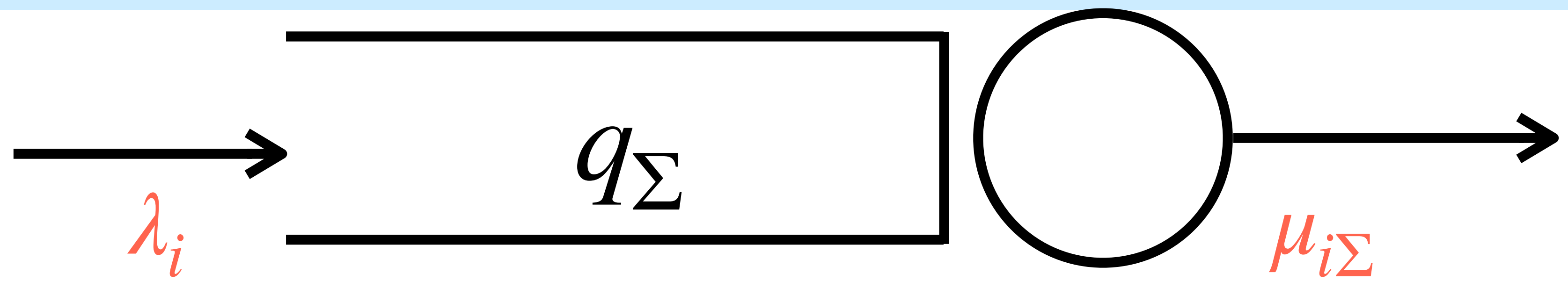
Key Takeaways

$\epsilon \boldsymbol{q} \sim ?$



$\approx$

Step 1: State Space Collapse



✓ Only need λ_i large enough so JSQ balances queue lengths for all i

✓ Queue can be unstable for some i

Step 2: Asymptotic distribution

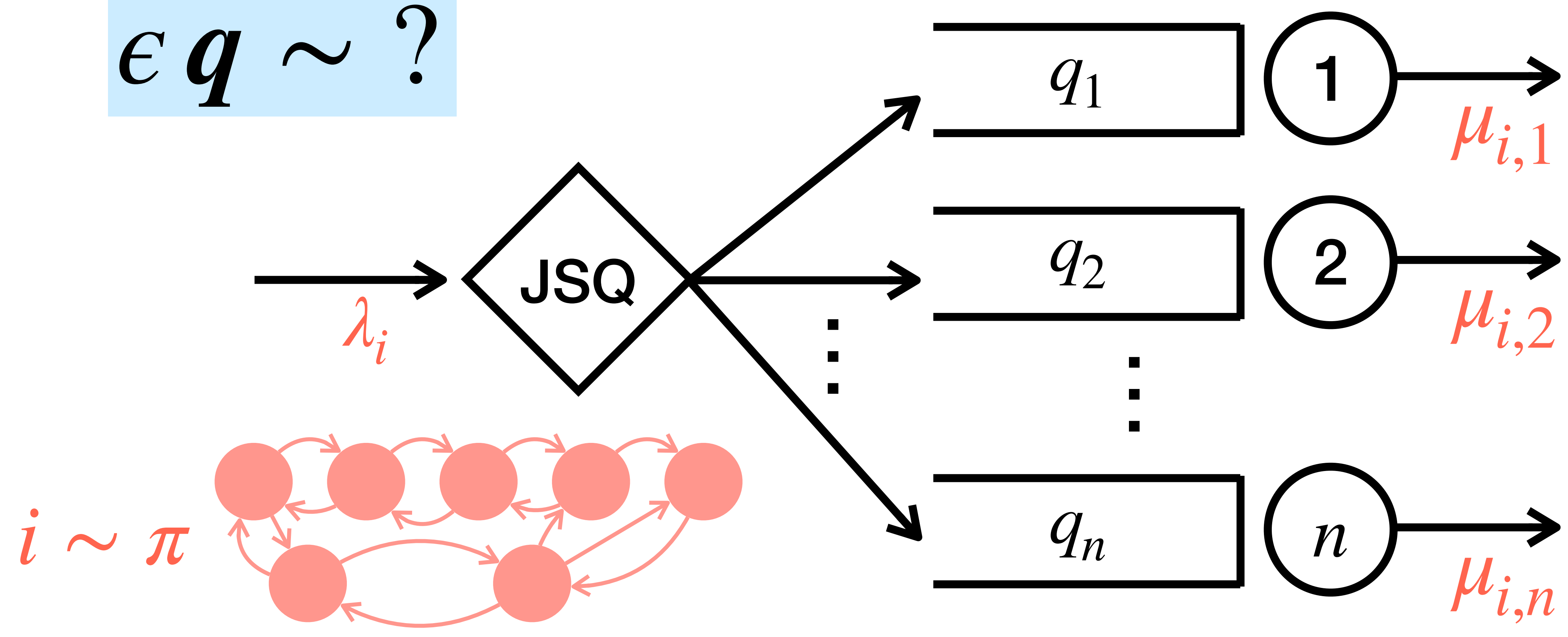
Transform Method:

$$\varphi_s(i, \boldsymbol{q}) = e^{-s\epsilon q_\Sigma}, \ \mathbb{E}[\Delta \varphi_s(i, \boldsymbol{q})] = 0$$

+ Poisson Equation

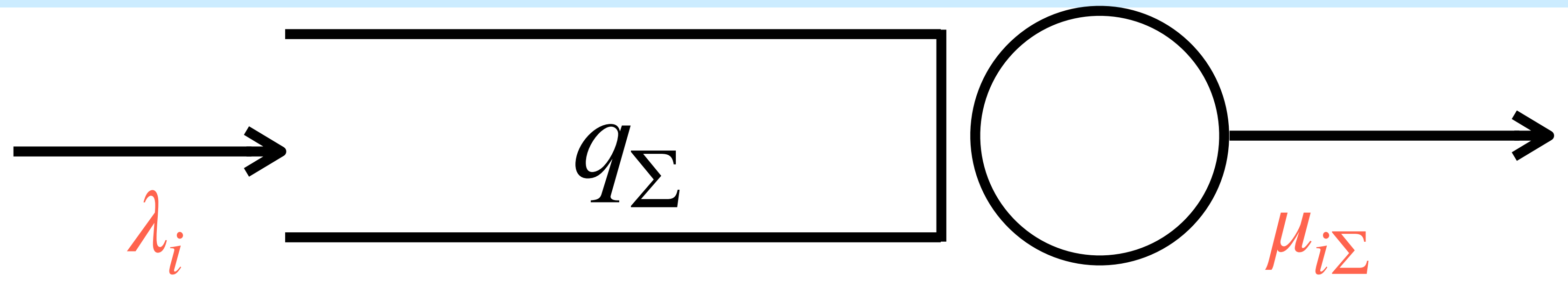
Key Takeaways

$\epsilon \boldsymbol{q} \sim ?$



$\approx$

Step 1: State Space Collapse



✓ Only need λ_i large enough so JSQ balances queue lengths for all i

✓ Queue can be unstable for some i

Theorem [HL, Grosof '25]:

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

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

Step 2: Asymptotic distribution

Transform Method:

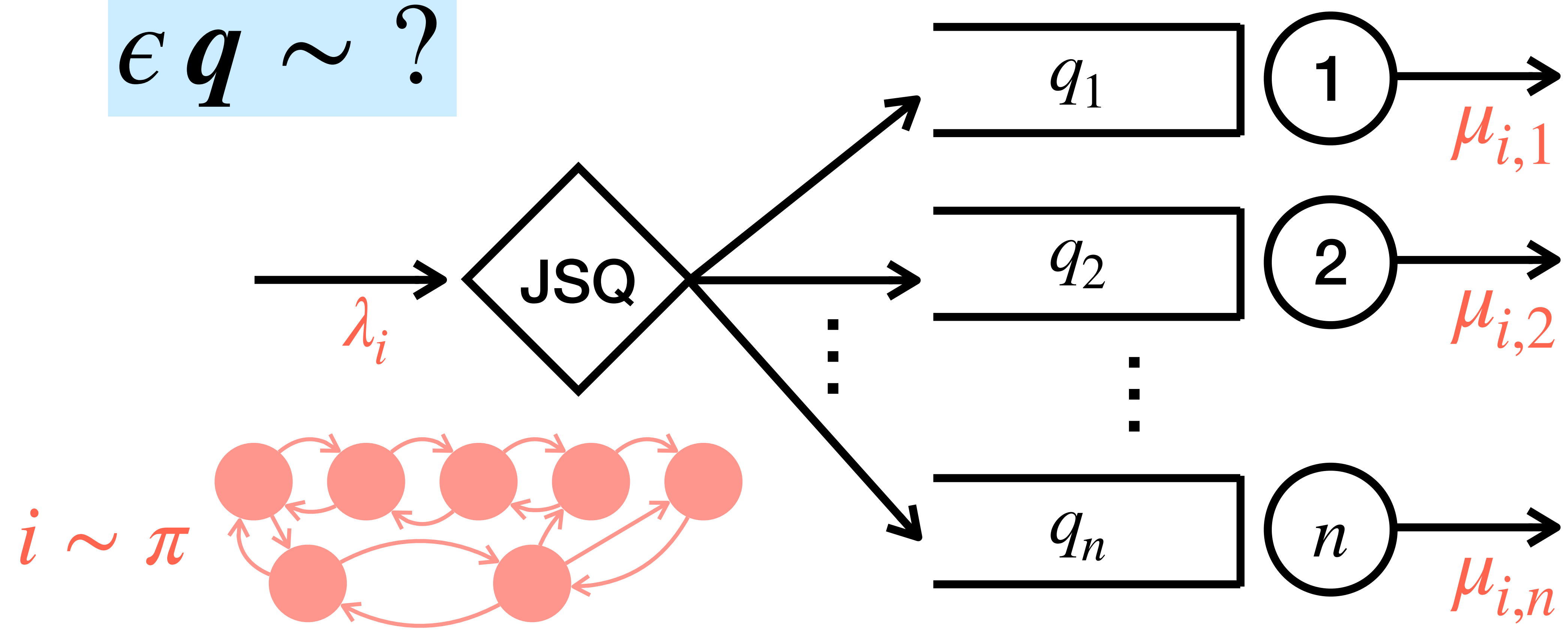
$$\varphi_s(i, \boldsymbol{q}) = e^{-s\epsilon q_\Sigma}, \ \mathbb{E}[\Delta \varphi_s(i, \boldsymbol{q})] = 0$$

+ Poisson Equation

Key Takeaways

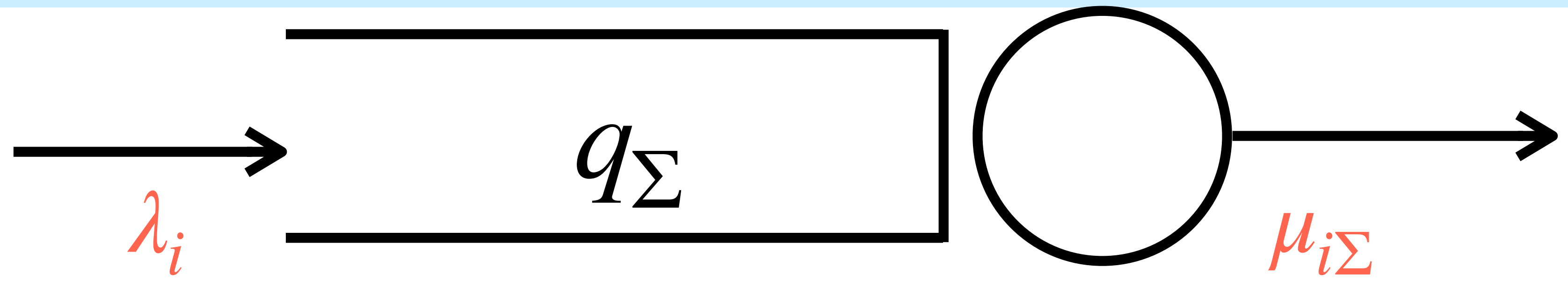
Thanks! Questions?

$\epsilon \boldsymbol{q} \sim ?$



$\approx$

Step 1: State Space Collapse



✓ Only need λ_i large enough so JSQ balances queue lengths for all i

✓ Queue can be unstable for some i

Theorem [HL, Grosof '25]:

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

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

Step 2: Asymptotic distribution

Transform Method:

$$\varphi_s(i, \boldsymbol{q}) = e^{-s\epsilon q_\Sigma}, \ \mathbb{E}[\Delta\varphi_s(i, \boldsymbol{q})] = 0$$

+ Poisson Equation