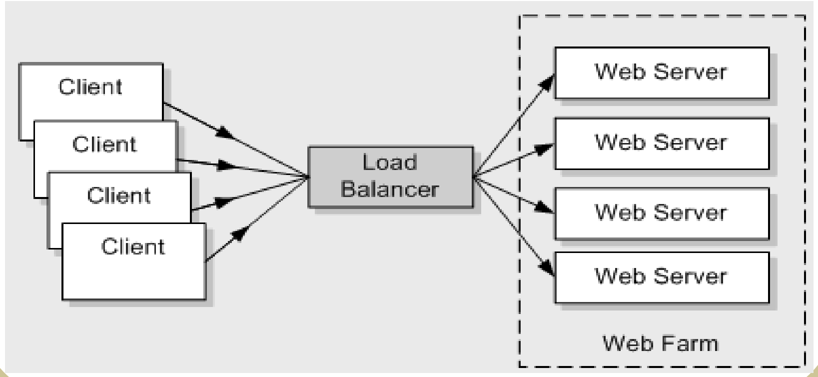
Approximations, Inapproximability and Tight Bounds for Queueing Systems

VARUN GUPTA
Carnegie Mellon University

Performance Evaluation and Design

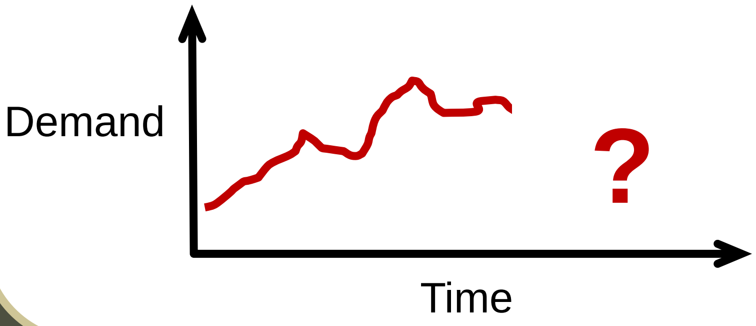
Scheduling algorithms

Load balancing algorithms

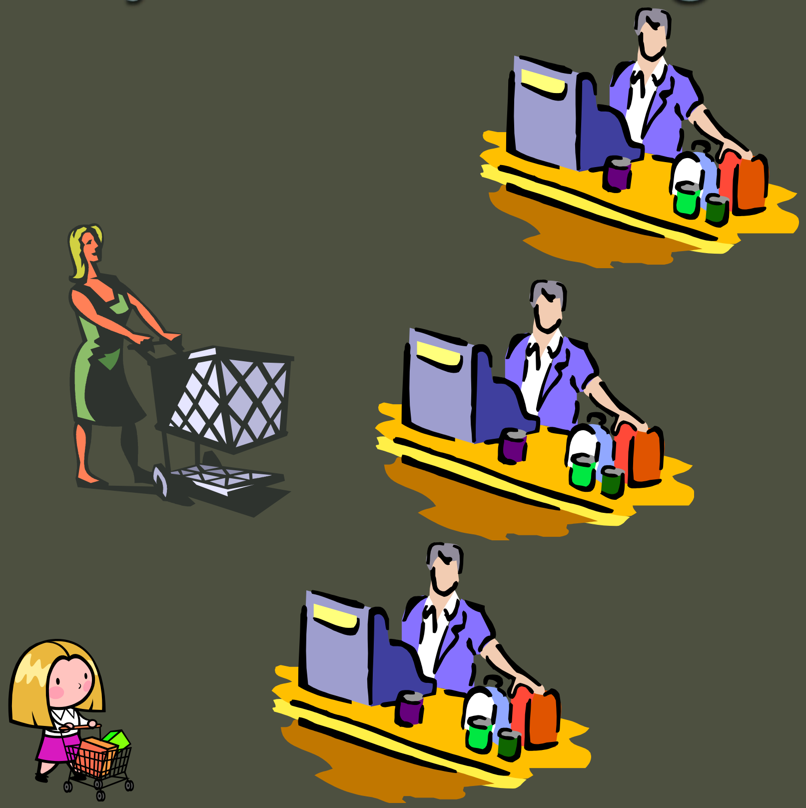


Capacity Provisioning (e.g., clouds, call centers,…)

Dynamic capacity scaling for energy-efficiency



Capacity Provisioning Questions



GOAL: Average Time in queue < t_{max}

Q: Minimum # open checkout counters?

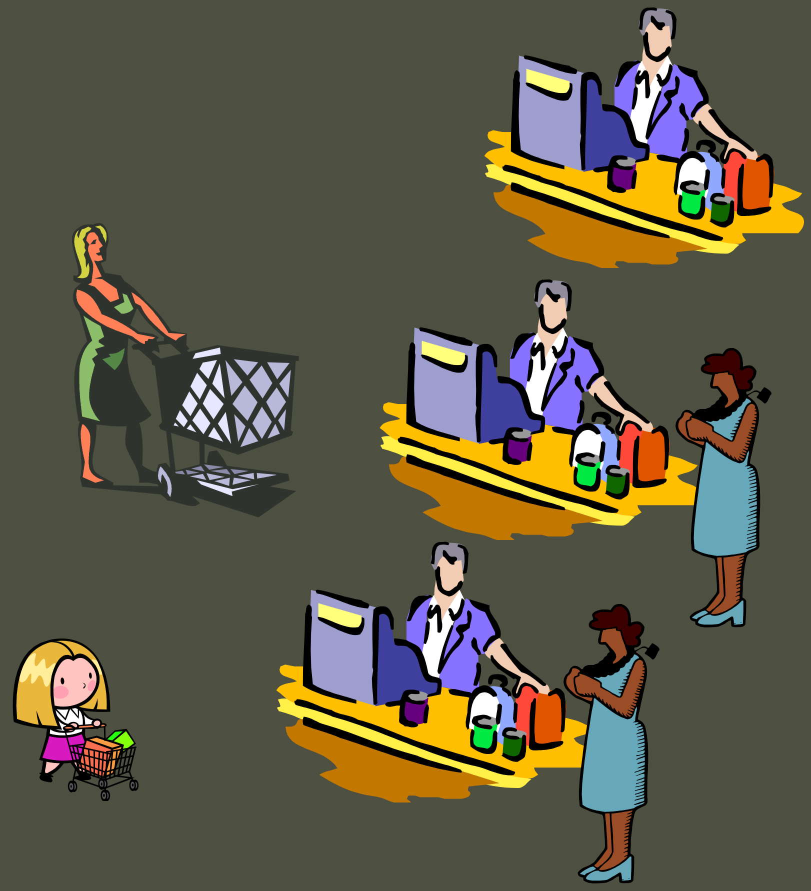
Capacity Provisioning Questions

GOAL: Average Time in queue < t_{max}

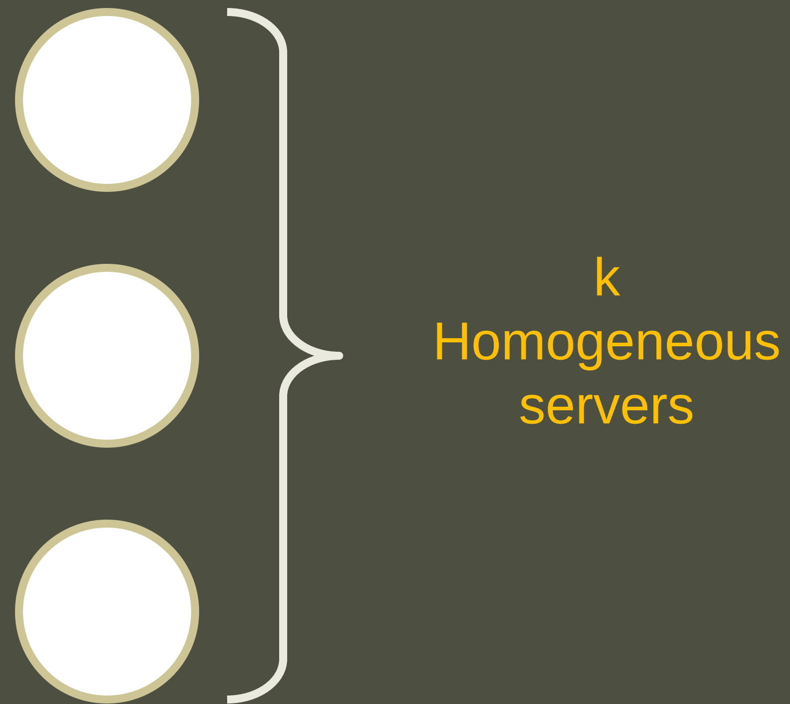
Q: Minimum # open checkout counters? 3 slow or 2 fast?

Stochastic Modeling (Queueing Theory) formalizes the above questions

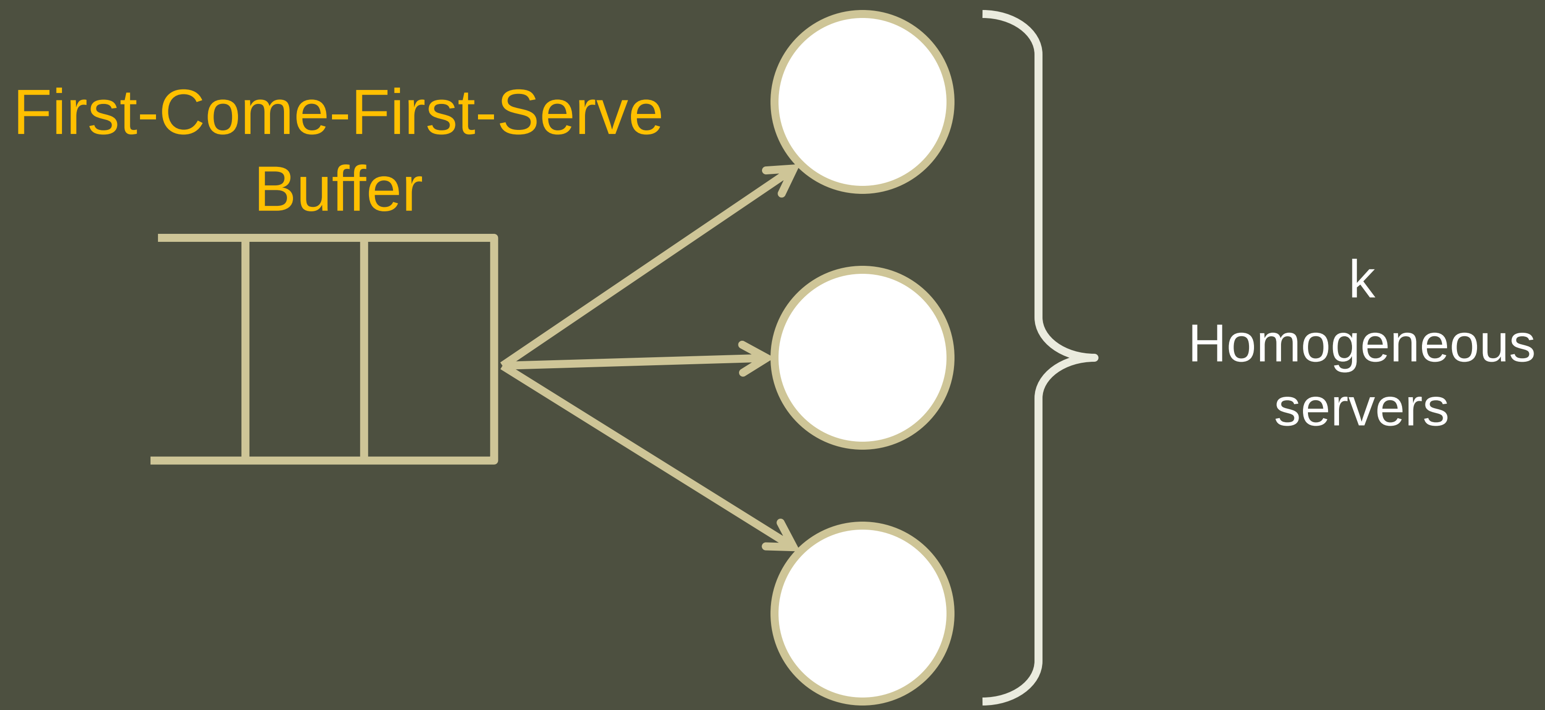
The *M/G/k/*FCFS model



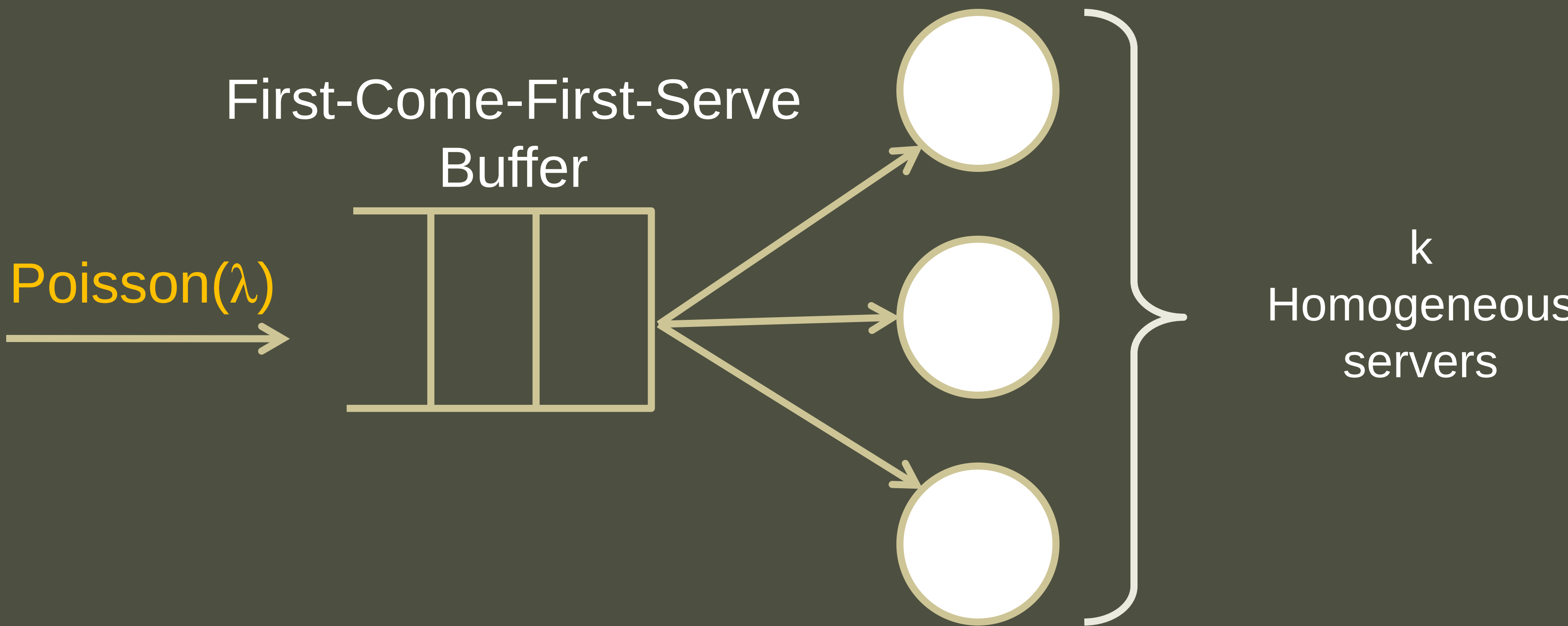
The *M/G/**k***/FCFS model



The *M/G/k/***FCFS** model

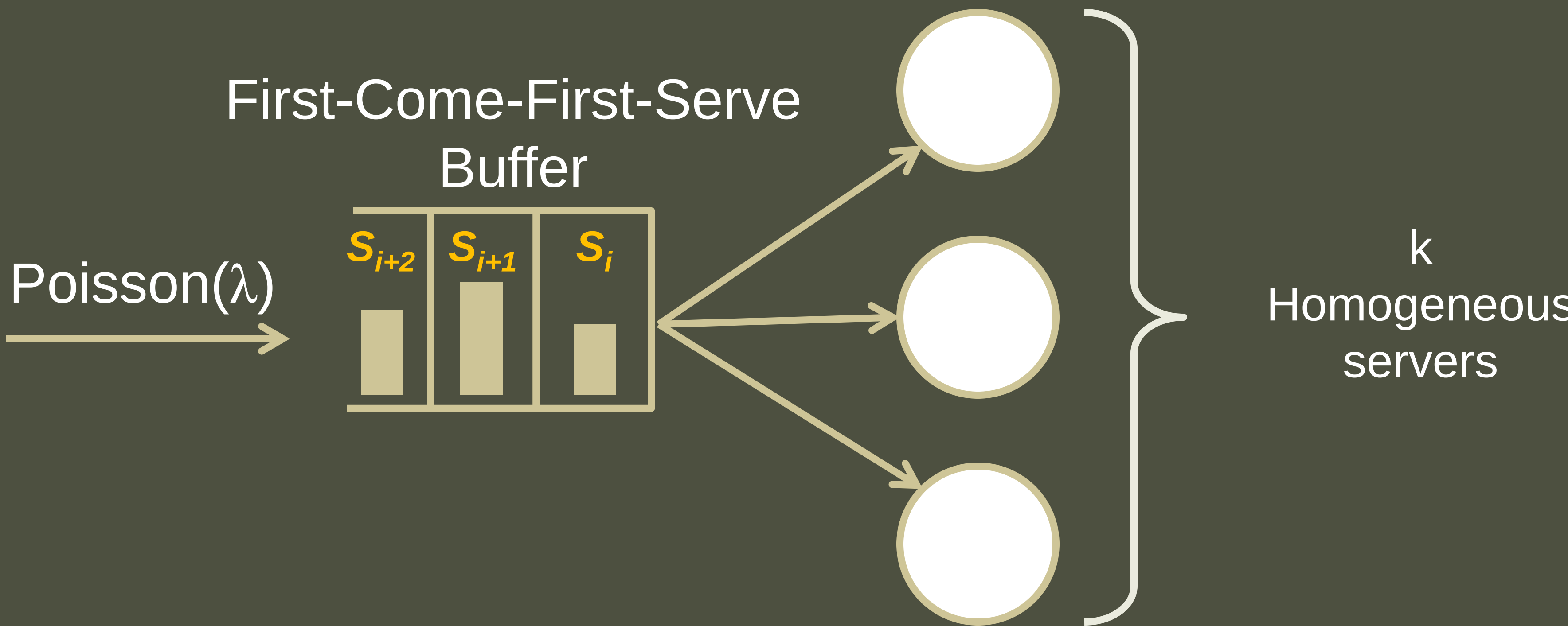


The *M*/*G*/*k*/FCFS model



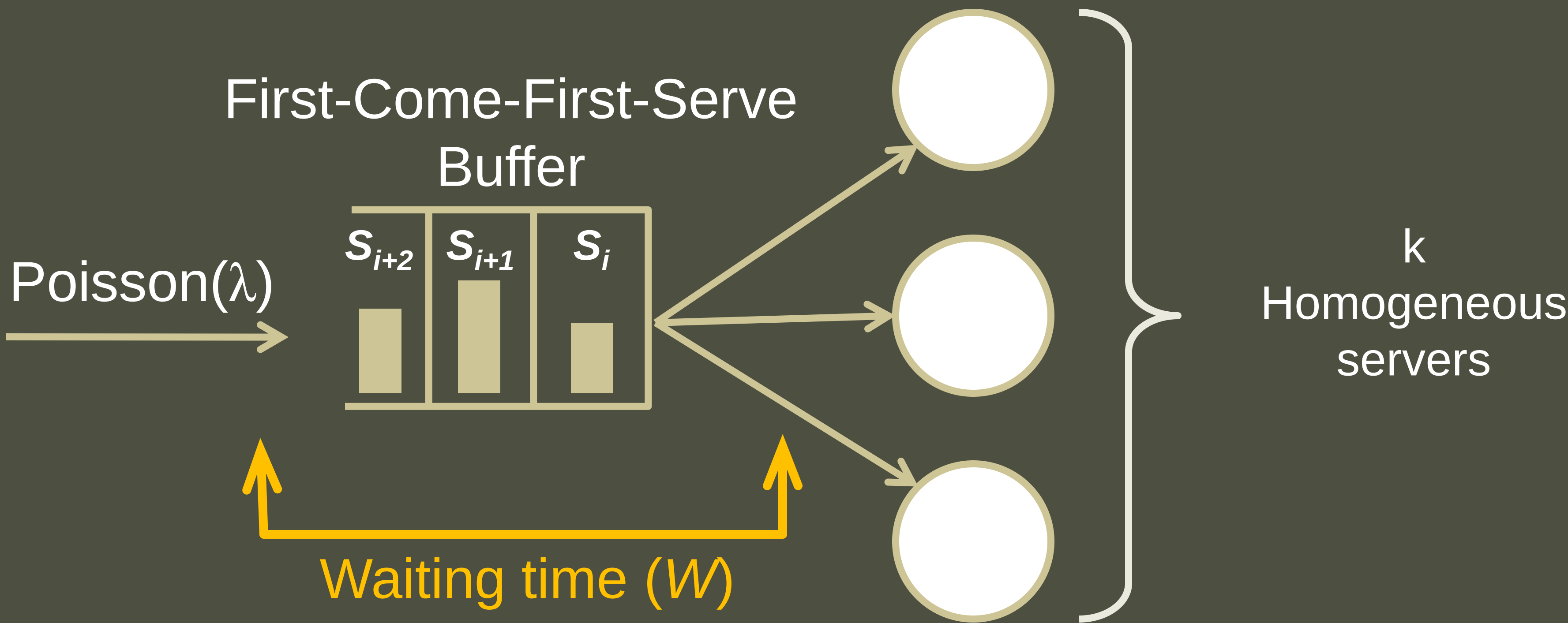
- λ = arrival rate

The *M*/*G*/*k*/FCFS model



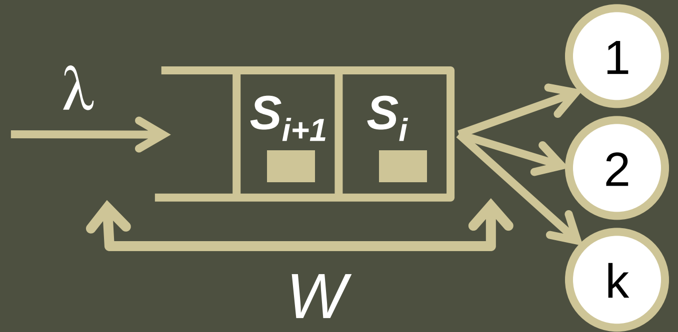
- λ = arrival rate
- job sizes (S_1, S_2, …) i.i.d. samples from S
- "load" $\rho \equiv \lambda\, E[S]$

The *M/G/k*/FCFS model



- λ = arrival rate
- job sizes (S_1, S_2, …) i.i.d. samples from S
- “load” $\rho \equiv \lambda\, E[S]$

GOAL : $E[W^{M/G/k}]$



$\rho \equiv \lambda\, E[S]$

k=1

Case : *S* ~ Exponential (*M/M/1*)
Analyze $E[W^{M/M/1}]$ via Markov chain (easy)

Case: *S* ~ General (*M/G/1*)

$$E[W^{M/G/1}] = \frac{C^2+1}{2} E[W^{M/M/1}]$$

$$C^2 = \frac{var(S)}{E[S]^2}$$

Sq. Coeff. of Variation (SCV)
> 20 for computing workloads

k>1

Case : S ~ Exponential (*M/M/k*)
$E[W^{M/M/k}]$ via Markov chain

Case: S ~ General (*M/G/k*)
No exact analysis known

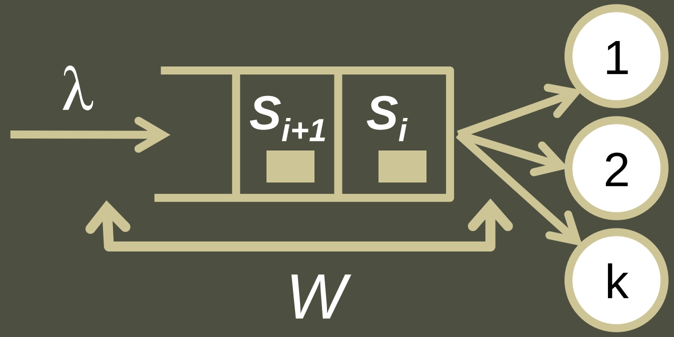
The Gold-standard approximation:

Lee, Longton (1959)

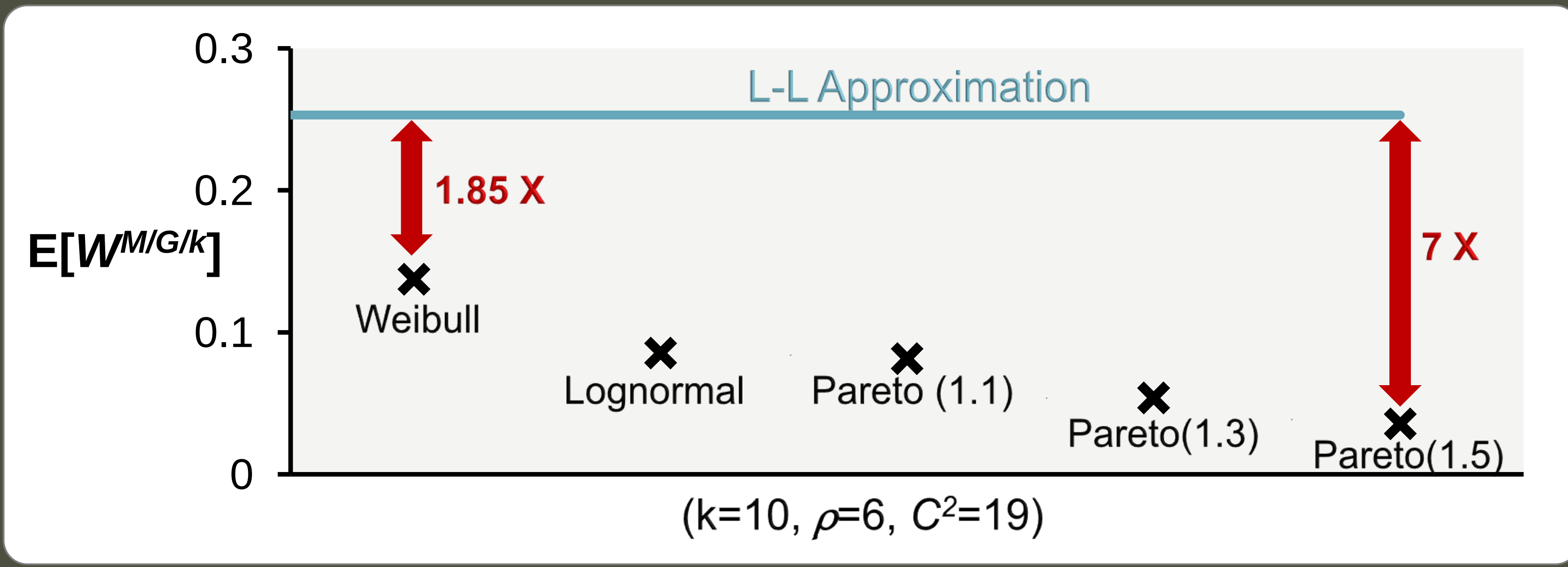
$$E[W^{M/G/k}] \approx \frac{C^2+1}{2} E[W^{M/M/k}]$$

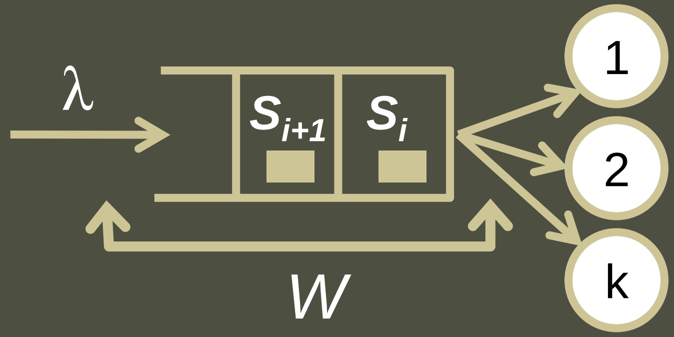
Lee, Longton approximation:

$$\mathrm{E}[W^{M/G/k}] \approx \frac{C^2+1}{2}\mathrm{E}[W^{M/M/k}]$$



- Simple
- Exact for k=1
- Asymptotically tight as $\rho \to k$ (think Central Limit Thm.)



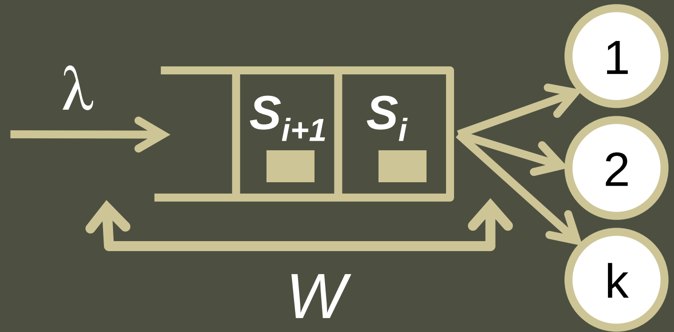


Outline

- An Inapproximability result for E[$W^{M/G/k}$]
- Framework for tight bounds via higher moments of S

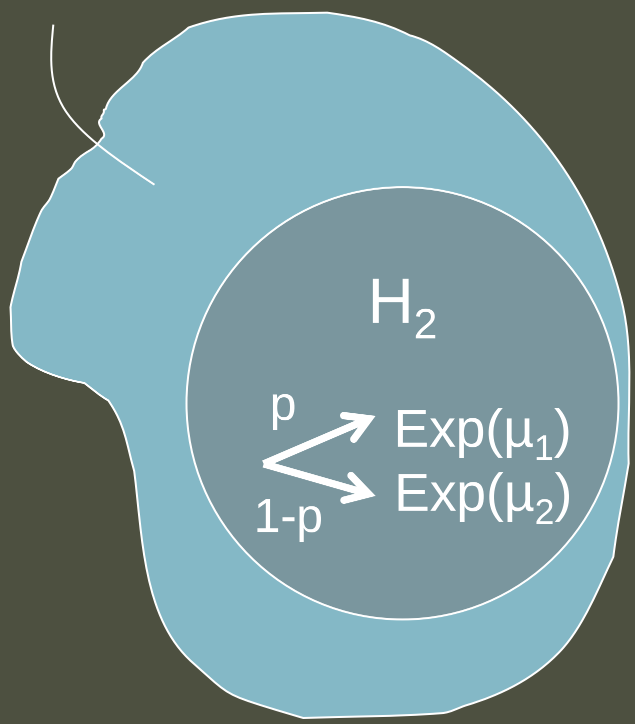
Lee Longton approximation:

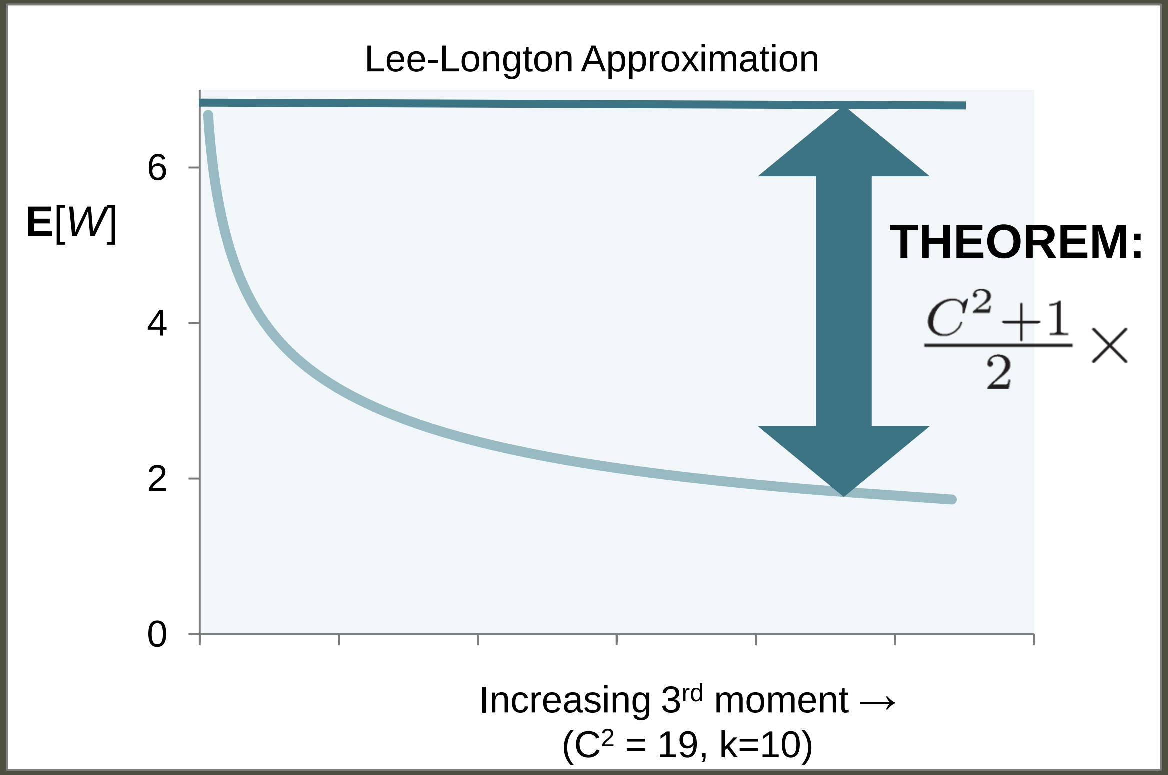
$$\mathrm{E}[W^{M/G/k}] \approx \frac{C^2+1}{2}\mathrm{E}[W^{M/M/k}]$$



GOAL: Bounds on approximation ratio

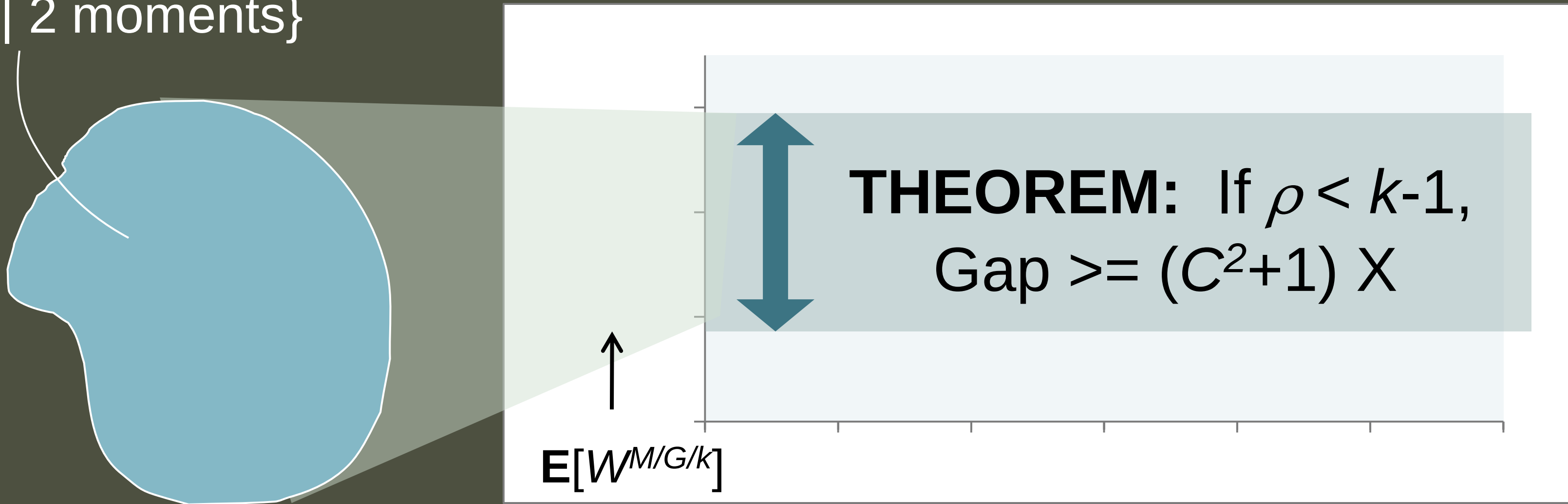
{G | 2 moments}





[Dai, G., Harchol-Balter, Zwart]

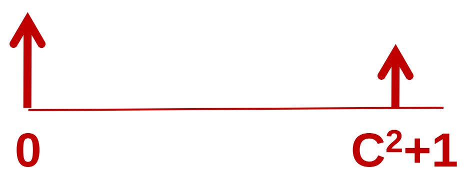
{G | 2 moments}



COR.: No approx. for E[$W^{M/G/k}$] based on first two moments of job sizes can be accurate for all distributions when C^2 is large

PROOF: Analyze limit distributions in D_2 ≡ mixture of 2 points

Min 3rd moment

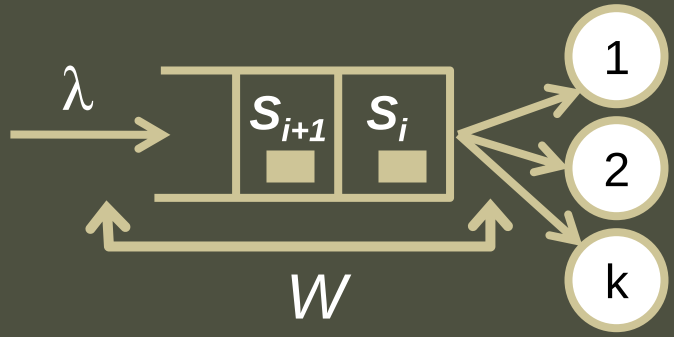


3rd moment → ∞

Approximations using higher moments?

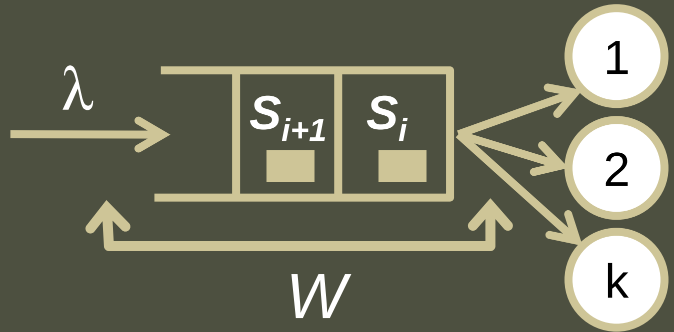
[Dai, G., Harchol-Balter, Zwart]

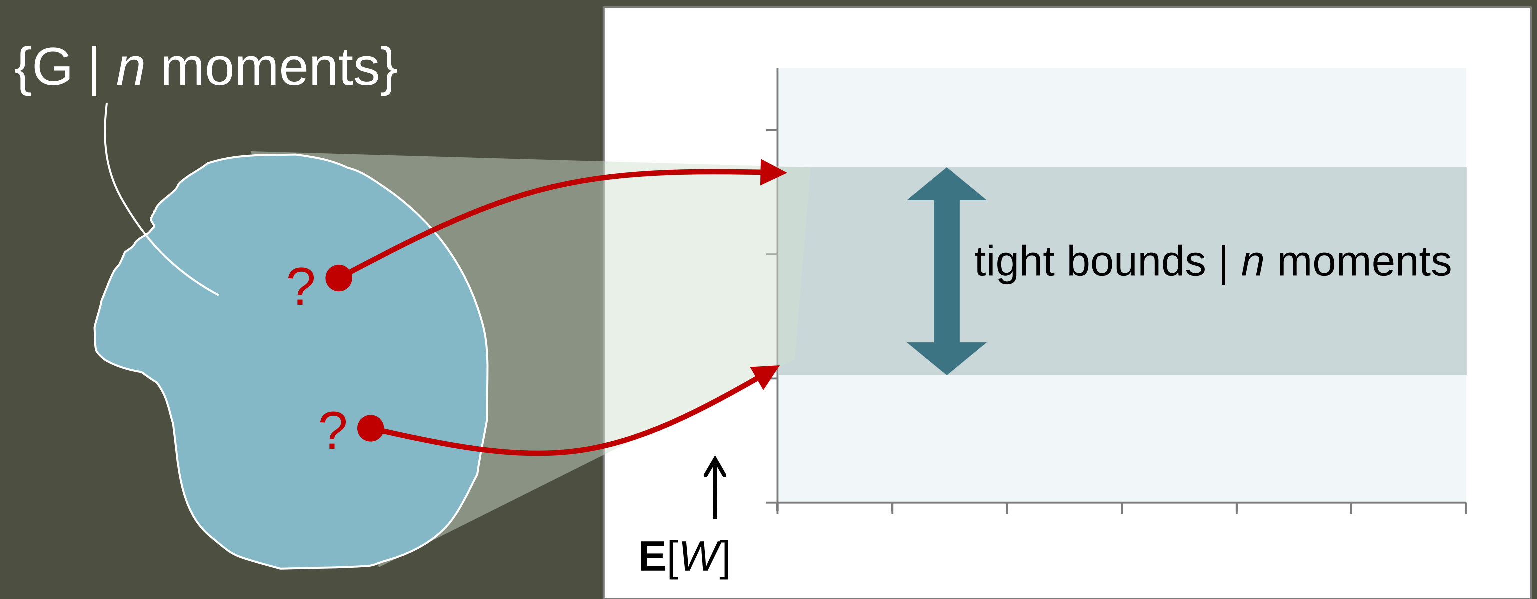
Outline



- An Inapproximability result for $E[W^{M/G/k}]$
- Framework for tight bounds via higher moments of S

Exploiting higher moments





GOAL: Identify the "extremal" distributions with given moments

RELAXED GOAL: Extremal distributions in some "non-trivial" asymptotic regime

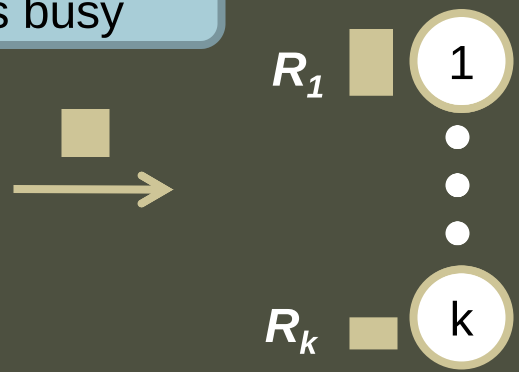
IDEA: Light-traffic asymptotics (λ→0)

RELAXATION: Identify the "extremal" distributions in light traffic

Light traffic theorem for *M/G/k* [Burman Smith]:

$$\mathrm{E}[W^{M/G/k}] = \frac{1}{k!}\left(\frac{\rho}{k}\right)^k \mathrm{E}[\min\{R_1, R_2, \ldots, R_k\}] + o(\rho^k)$$

Probability of finding all servers busy

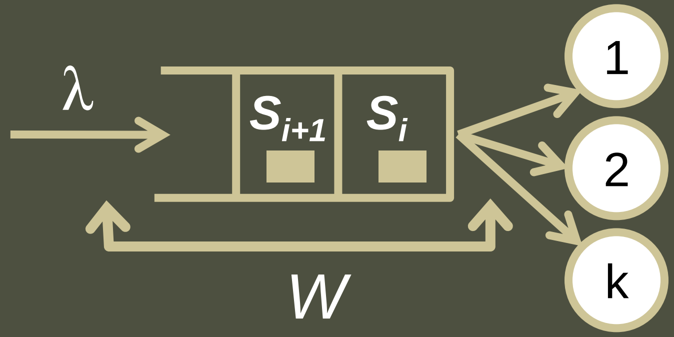


i.i.d. copies of *R* ≡ *equilibrium residual size* of *S*

pdf of *R*: $f_R(x) = \frac{\mathrm{Prob}[S \geq x]}{\mathrm{E}[S]}$

SUBGOAL: Extremal distributions for E[min{R_1,…,R_k}]
s.t. E[S^i] = m_i for i=1,..,n

Where we are…



GOAL: Tight bounds on E[$W^{M/G/k}$] given *n* moments of *S*

IDEA: Identify extremal distributions

RELAXATION (Light Traffic): Extremal distributions for

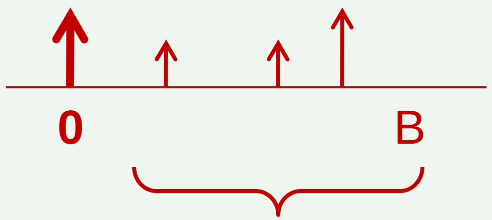
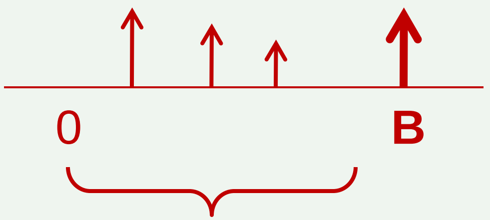
E[min{R_1,…,R_k}] s.t. E[S^i] = m_i for i=1,..,*n*

Principal Representations, Extremal Problems, and Tchebycheff-systems

GIVEN: Moment conditions on random variable X with support [0,B]

$$
\begin{gathered}
E[f_0(X)]=m_0 \\
E[f_1(X)]=m_1 \\
\cdots \\
E[f_n(X)]=m_n
\end{gathered}
$$

Principal Representations (p.r.) on [0,B] are distributions satisfying the moment conditions, and the following constraints on the support

	Lower p.r.	Upper p.r.
n even	 **1** + n/2 point masses	 **1** + n/2 point masses

Principal Representations, Extremal Problems, and Tchebycheff-systems

GIVEN: Moment conditions on random variable X with support [0,B]

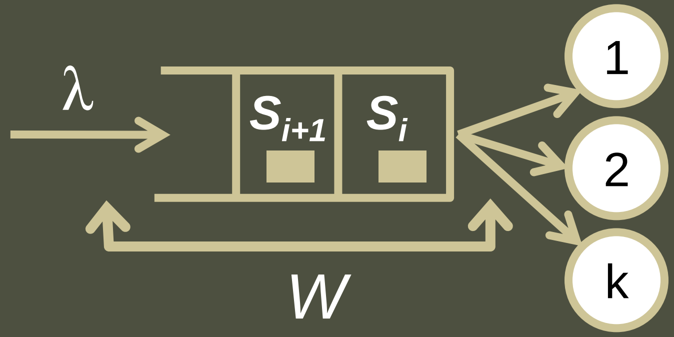
$$E[f_0(X)]=m_0$$
$$E[f_1(X)]=m_1$$
$$\ldots$$
$$E[f_n(X)]=m_n$$

Want to bound: E[g(X)]

THEOREM [Markov-Krein]:

If $\{f_0,f_1,\ldots,f_n\}$ and $\{f_0,\ldots,f_n,g\}$ are Tchebycheff-systems on [0,B], then E[g(X)] is extremized by the unique lower and upper principal representations of the moment sequence $\{m_0,\ldots,m_n\}$.

Where we are…



GOAL: Tight bounds on E[$W^{M/G/k}$] given n moments of S

IDEA: Identify extremal distributions

RELAXATION (Light Traffic): Extremal distributions for

$E[\min\{R_1,\ldots,R_k\}]$ s.t. $E[S^i] = m_i$ for i=1,..,n

RELAXATION: Extremal distributions for $E[\min\{R_1,\ldots,R_k\}]$
s.t. $E[S^i] = m_i$ for i=1,..,n

IDEA 1: Want to use Markov-Krein Theorem to say upper/ lower p.r. of $\{m_0, \ldots, m_n\}$ are extremal …

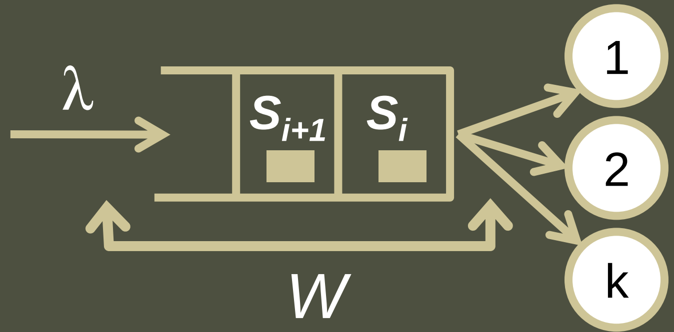
IDEA 2: Suffices to prove p.r.s extremize

$$E[\min\{R_1,c_2,\ldots,c_k\}] = E[g(S)] \text{ for all } c_2,\ldots,c_k > 0$$

g(x) piecewise polynomial, but does not form a T-system with x^i

THEOREM [G., Osogami]: Upper and lower p.r. are extremal for $E[\min\{R_1,\ldots,R_k\}]$
s.t. $E[S^i] = m_i$ for i=1,..,n, if n=2 or 3.

Where we are…



GOAL: Tight bounds on $E[W^{M/G/k}]$ given n moments of S

IDEA: Identify extremal distributions

RELAXATION (Light Traffic): Extremal distributions for

$E[\min\{R_1,\ldots,R_k\}]$ s.t. $E[S^i] = m_i$ for i=1,..,n

THEOREM:
For n = 2 or 3

RELAXATION 2: Restrict to mixtures of Exponential distributions

(Dense in Completely Monotone (CM) family;
CM contains Weibull, Pareto, Gamma)

THEOREM:
For all n.

SUBGOAL: Extremal distributions for E[min{R_1,…,R_k}]
s.t. E[S^i] = m_i for i=1,..,n; and S is mixture of Exponential

IDEA 1: Want to use Markov-Krein Theorem to say upper/ lower p.r. of {m_0, …,m_n} *within this class* are extremal …

Need to define upper/lower p.r. for mixtures of Exponentials

S ≡ mixture of Exponential distributions ⟺ Y_S ≡ mixture of point masses

Moments of S ⟹ Moments of Y_S

⇓ ⇓

p.r. of S ⟸ p.r. of Y_S

SUBGOAL: Extremal distributions for $E[\min\{R_1,\ldots,R_k\}]$
s.t. $E[S^i] = m_i$ for i=1,..,n; and S is mixture of Exponential

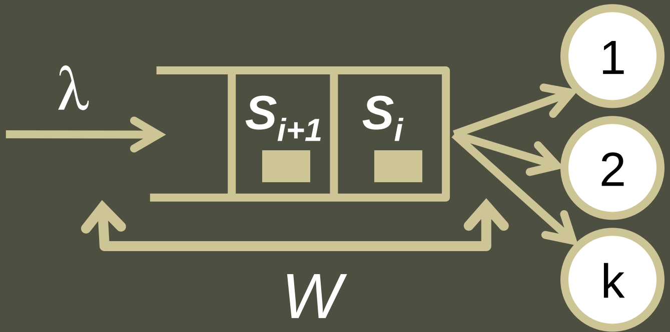
IDEA 1: Want to use Markov-Krein Theorem to say upper/ lower p.r. of $\{m_0, \ldots, m_n\}$ *within this class* are extremal …

IDEA 2: Suffices to prove p.r.s extremize
$E[\min\{R_1, \mathrm{Exp}(c_2), \ldots, \mathrm{Exp}(c_k)\}] = E[g(Y_S)]$ for all $c_2,\ldots,c_k > 0$

$g(y) = a+b/(cy+1)$, and does form a T-system with y^i

THEOREM [G., Osogami]: Upper and lower p.r. are extremal for $E[\min\{R_1,\ldots,R_k\}]$
s.t. $E[S^i] = m_i$ for i=1,..,n; and S mixture of Exponential, $\forall$n.

Where we are…



GOAL: Tight bounds on $E[W^{M/G/k}]$ given n moments of S
IDEA: Identify extremal distributions

RELAXATION (Light Traffic): Extremal distributions for

$E[\min\{R_1,\dots,R_k\}]$ s.t. $E[S^i] = m_i$ for $i=1,..,n$

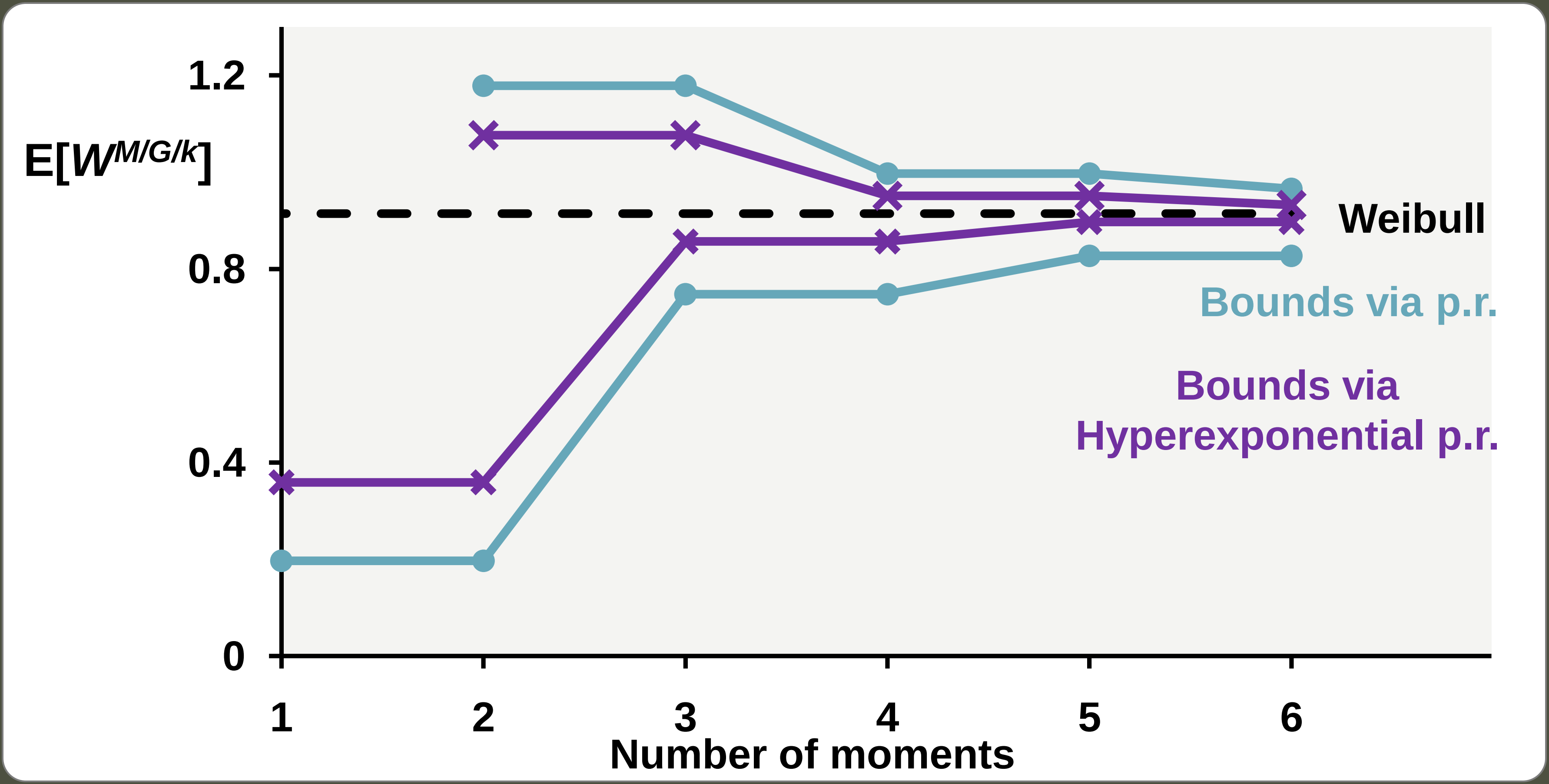
THEOREM:
For n = 2 or 3

RELAXATION 2: Restrict to mixtures of Exponential distributions

(Dense in Completely Monotone (CM) family;
CM contains Weibull, Pareto, Gamma)

THEOREM:
For all n.

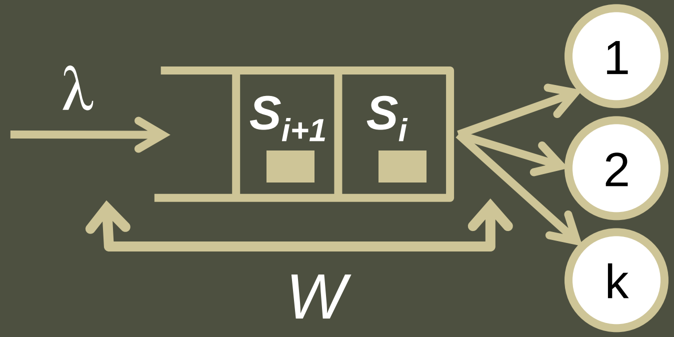
Simulation Results (k=4, ρ=2.4,)



Approximation Schema:

Refine lower bound via an additional odd moment,

Upper bound via even moment until gap is acceptable

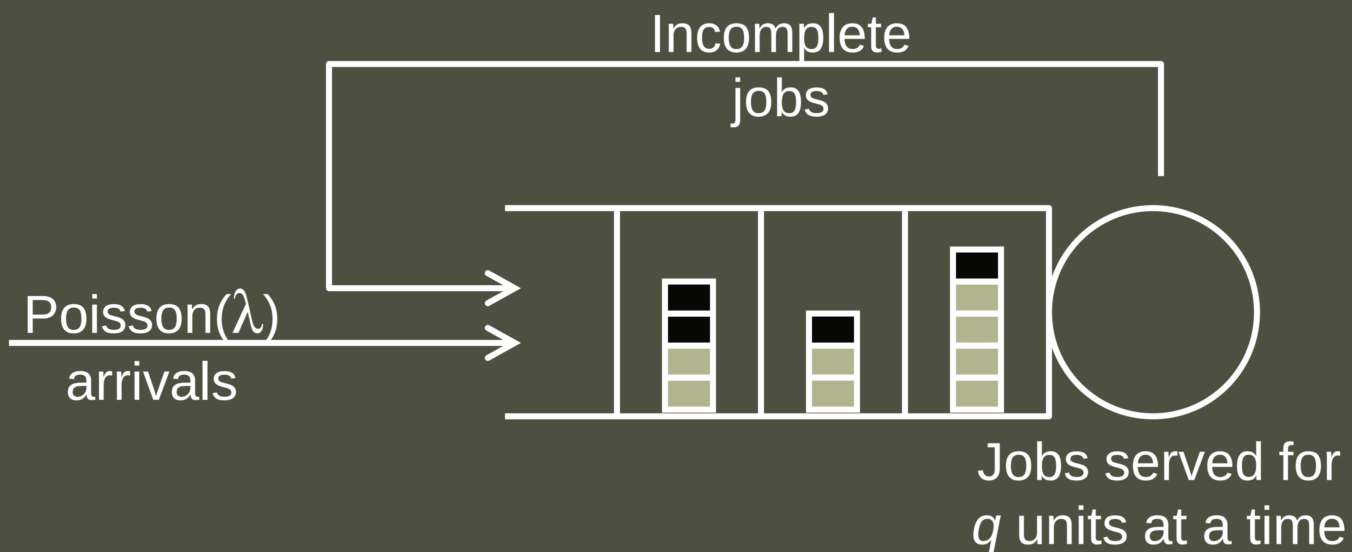


Outline

- An inapproximability result for $E[W^{M/G/k}]$
- Framework for tight bounds via higher moments of S
- Many other “hard” queuing systems fit the above framework too

Other queuing systems exhibiting Markov-Krein characterization

Example 1: M/G/1 Round-robin queue

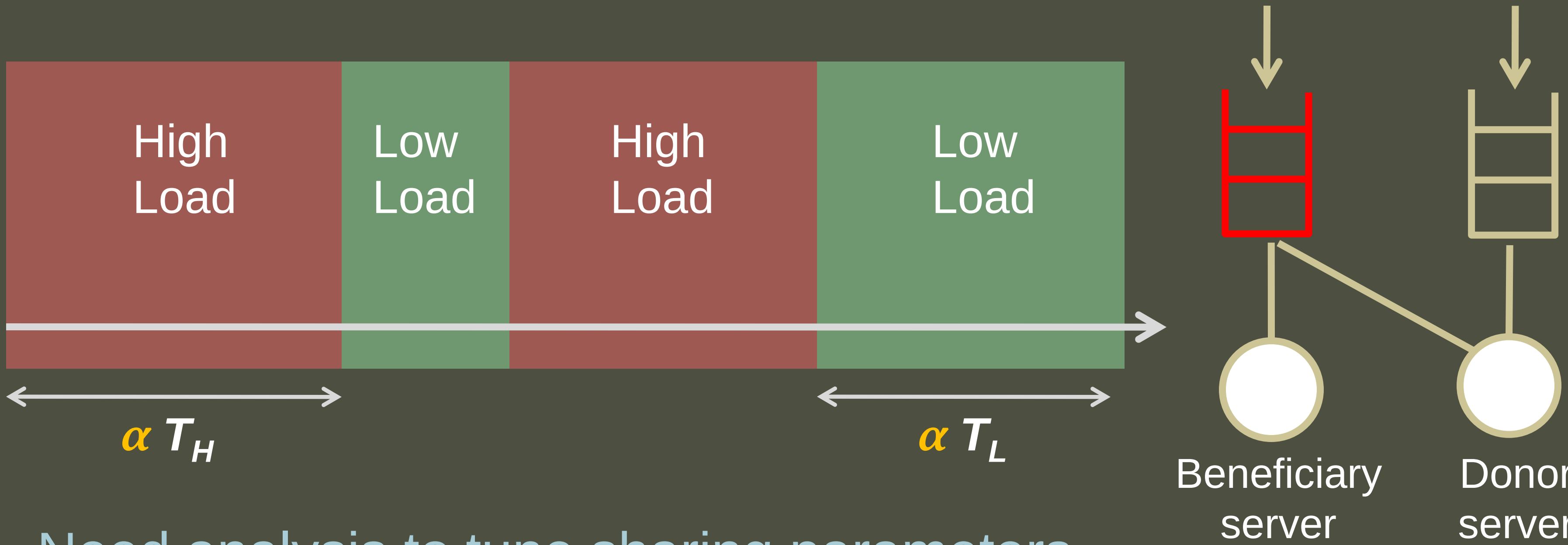


Need analysis to find *q* that balance overheads/performance

THEOREM [G., Osogami]: Upper and lower p.r. extremize mean waiting time under $\lambda \to 0$, when *S* is mixture of Exponential.

Other queuing systems exhibiting Markov-Krein characterization

Example 2: Systems with fluctuating load



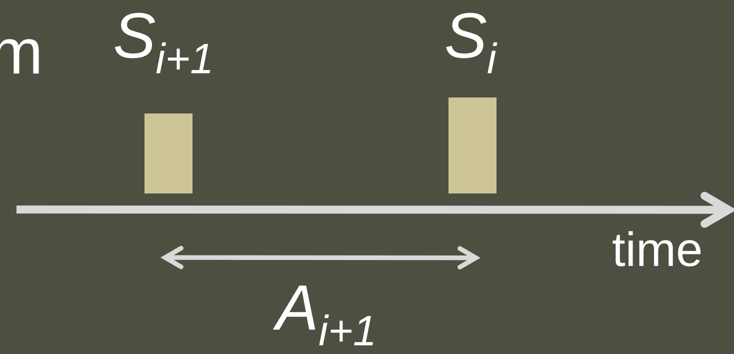
Need analysis to tune sharing parameters

THEOREM [G., Osogami]: Upper and lower p.r. extremize mean waiting time under $\alpha \to 0$, when T_H, T_L are mixtures of Exponential.

Open problem: Markov-Krein characterization of Stochastic Recursive Sequences

Example: Single server FCFS system

W_{i+1} = waiting time of S_{i+1}

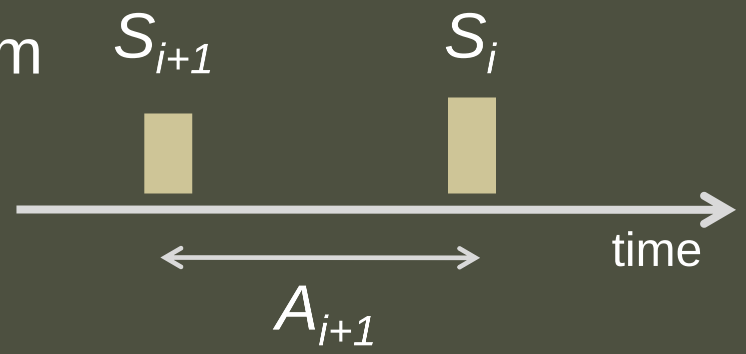


$$W_{i+1} = \Phi(W_i, S_i, A_{i+1})$$

Open problem: Markov-Krein characterization of Stochastic Recursive Sequences

Example: Single server FCFS system

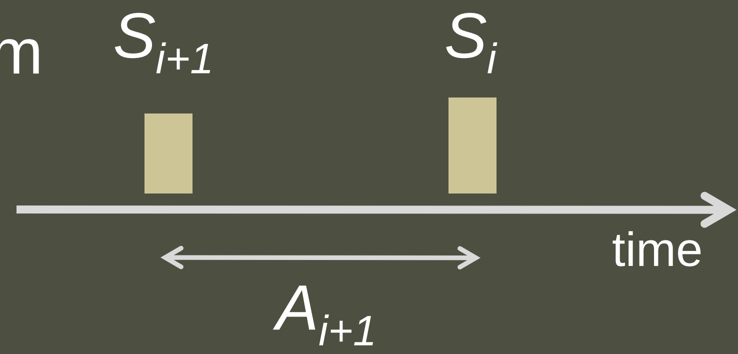
W_{i+1} = waiting time of S_{i+1}



$$W_{i+1} = (W_i + S_i - A_{i+1})^+$$

Open problem: Markov-Krein characterization of Stochastic Recursive Sequences

Example: Single server FCFS system



W_{i+1} = waiting time of S_{i+1}

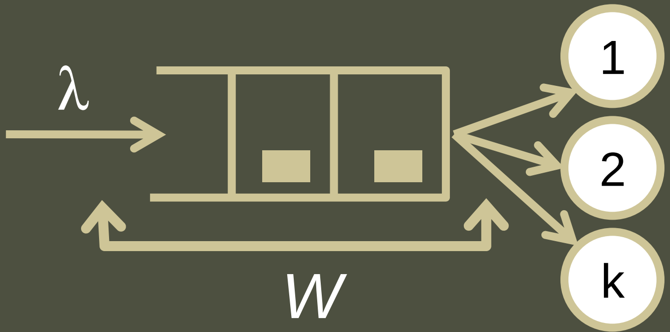
$$W \overset{d}{=} (W + S - A)^+$$

Stationary behavior of a queueing system = Fixed point of a stochastic recursive sequence of the form

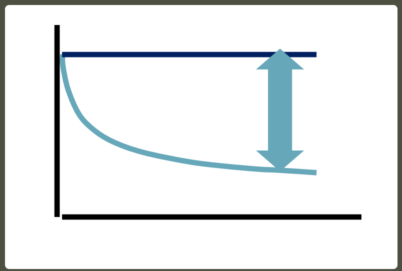
$$W \overset{d}{=} \Phi(W,S)$$

Q: Given moments of *S*, under what conditions on f, Φ, is E[f(*W*)] extremized by p.r.s?

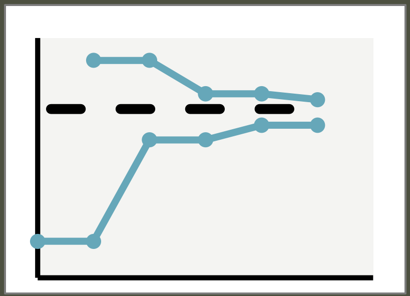
Conclusions



- All existing analytical approx for performance based on 2 moments, but 2 moments inadequate



- Provide evidence for tight *n*-moments based bounds via asymptotics for *M/G/k* and other queuing systems



- A new problem in analysis: Markov-Krein characterization of stochastic fixed point equations

$$W \overset{d}{=} \Phi(W,S)$$

THEOREM [Markov-Krein]:

If $\{f_0, f_1, \ldots, f_n\}$ and $\{f_0, \ldots, f_n, g\}$ are Tchebycheff-systems on [0,B], then E[$g(X)$] is extremized by the unique lower and upper principal representations of the moment sequence $\{m_0, \ldots, m_n\}$.

Tchebycheff-system

$\{f_0, f_1, \ldots, f_n\}$ form a Tchebycheff-system on [0,B] if

$$a_0 f_0 + a_1 f_1 + \ldots + a_n f_n$$

has <= n roots (counting multiplicities) in [0,B] for any $a_0, a_1, \ldots, a_n$

Example 1 (Power functions): $f_i(x) = x^i$

Example 2 (Cauchy kernel): $f_i(x) = 1/(c_i + x)$ for $c_i > 0$