Characterization of the output process for some fluid queues

Samuli Aalto
EURANDOM
Eindhoven

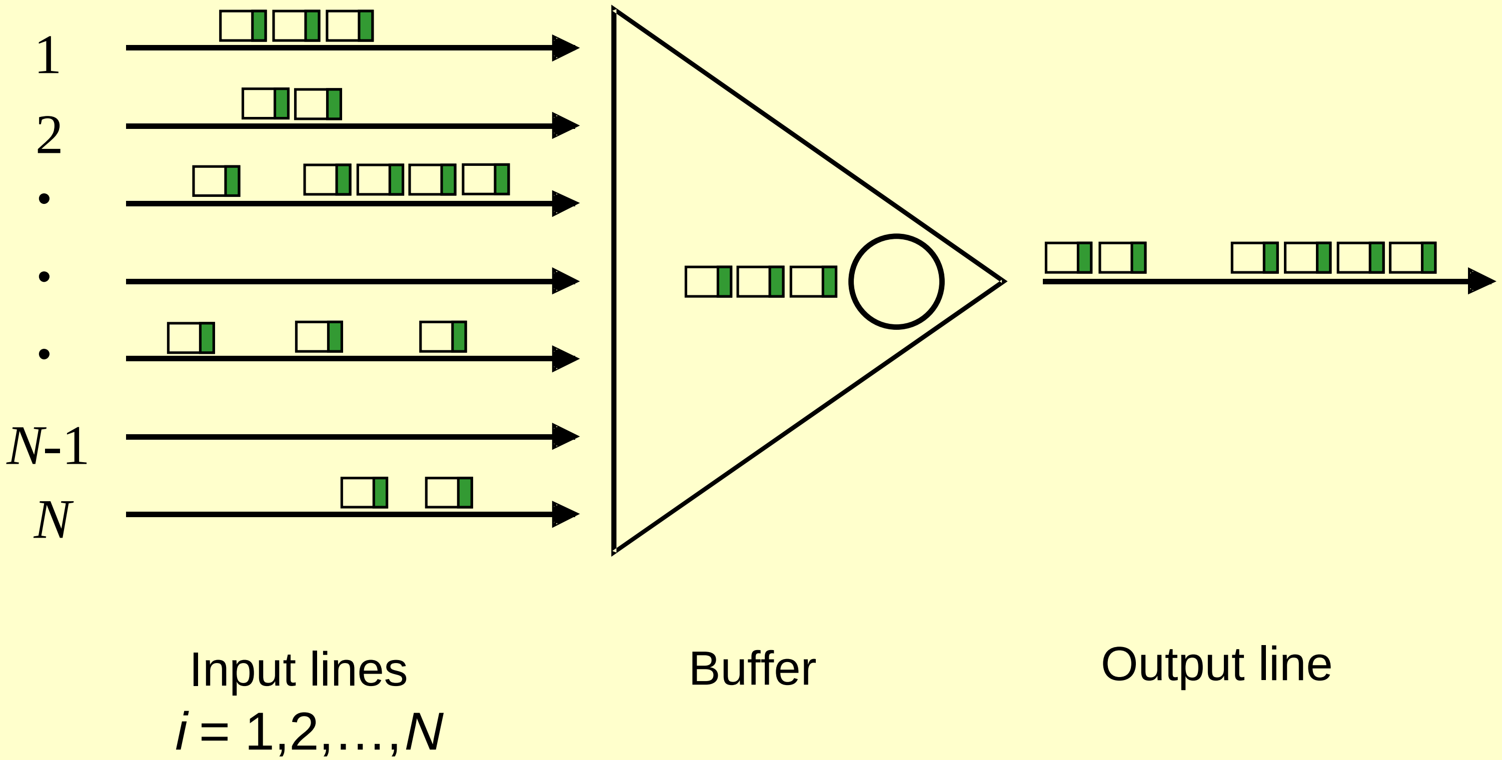
Background

- **Ph.D. Thesis**: “Studies in Queueing Theory”, University of Helsinki, Finland, 1998
  - [1] S. Aalto (1997) Optimal control of batch service queues with Poisson arrivals and finite service capacity, Dep Mathematics, University of Helsinki
  - [2] S. Aalto (1998) Optimal control of batch service queues with compound Poisson arrivals and finite service capacity, *Math Meth Oper Res*
  - **[3]** S. Aalto (1998) Characterization of the output rate process for a Markovian storage model, *J Appl Prob*
  - **[4]** S. Aalto (1998) Output of a multiplexer loaded by heterogeneous on-off sources, *Stoch Models*
- Supervisors: Prof. E. Nummelin and Ph.D. T. Lehtonen

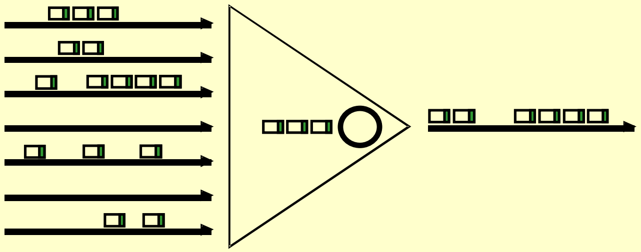
Contents

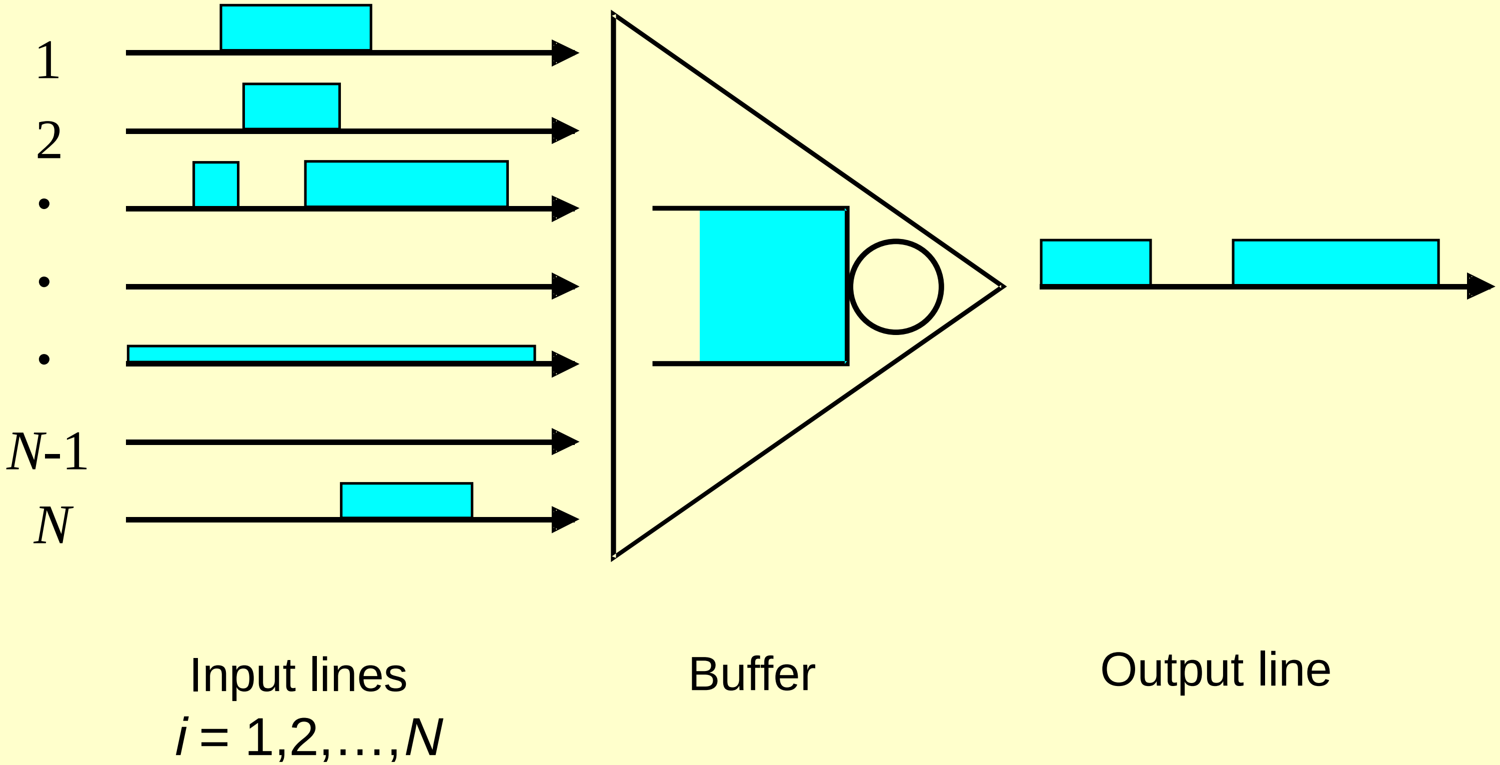
- Fluid queue
- Fluid queues driven by a single on-off source
- Fluid queues driven by multiple on-off sources
- Fluid queues driven by Markov jump processes
- Tandem fluid queues

Statistical multiplexer ...

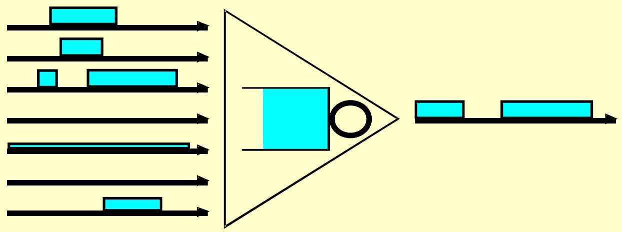


... as seen at burst level





Fluid queue = fluid flow storage model



- Input rate $r_0(t)$
  - varying randomly
  - gradual input!
- Buffer size
  - finite or infinite
- Leak rate c_1
  - max output rate
  - gradual output!

- Buffer content process $Z(t)$
- Output rate process $r_1(t)$

$$Z(t) = Z(0) + \int_0^t r_0(u)\,du - \int_0^t r_1(u)\,du$$

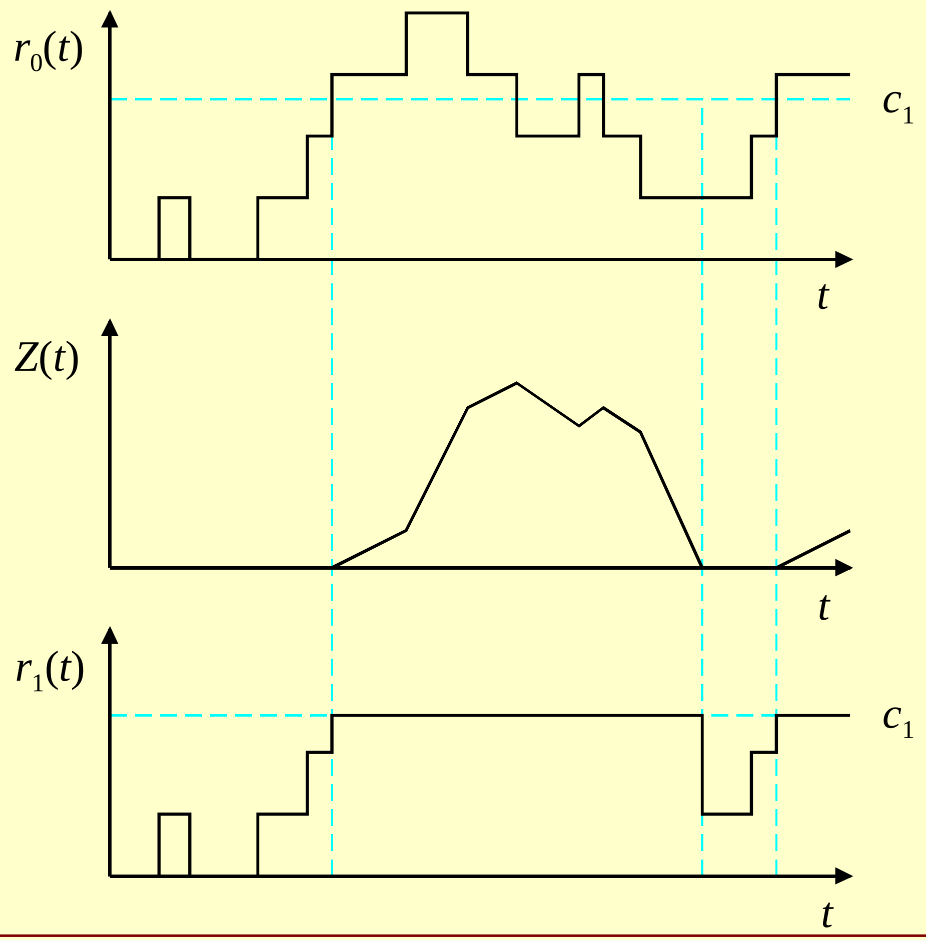
$$r_1(t) = \begin{cases} \min\{r_0(t), c_1\}, & \text{if } Z(t) = 0 \\ c_1, & \text{if } Z(t) > 0 \end{cases}$$

Evolution

Input rate $r_0(t)$

Buffer content $Z(t)$

Output rate $r_1(t)$



Leak rate c_1

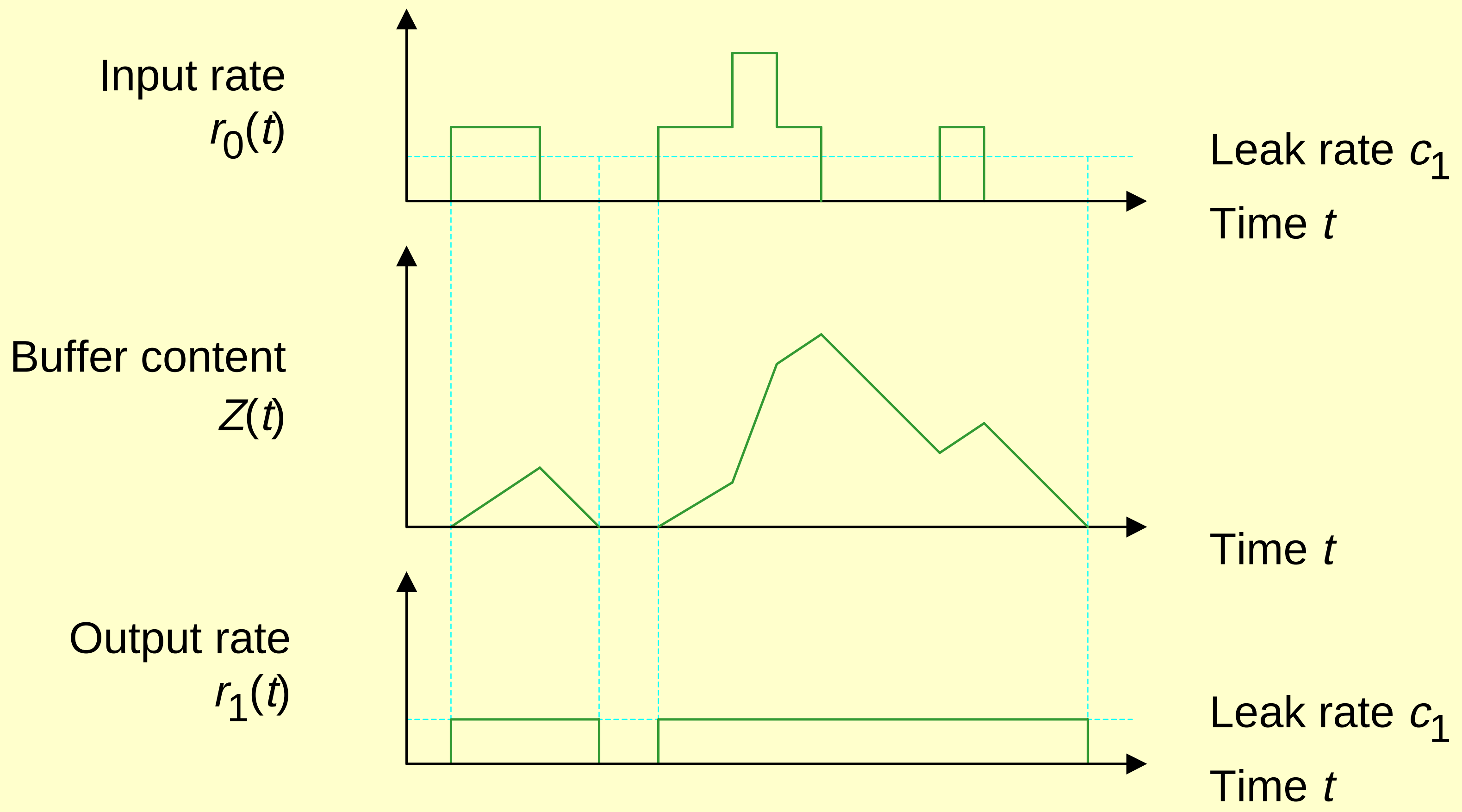
Time t

Time t

Leak rate c_1

Time t

Evolution assumed that $r_0(t) > c_1$ whenever $r_0(t) > 0$

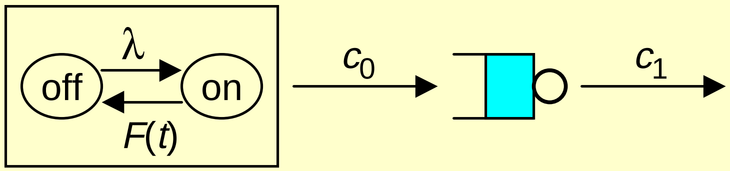


Contents

- Fluid queue
- Fluid queues driven by a single on-off source
- Fluid queues driven by multiple on-off sources
- Fluid queues driven by Markov jump processes
- Tandem fluid queues

Fluid queue driven by a single on-off source

- On-off source with rate c_0
  - Idle periods ~ Exp(λ)
  - Active periods ~ $F(t)$



- $c_0 \leq c_1$ => input = output
- Assumption:

$$c_0 > c_1$$

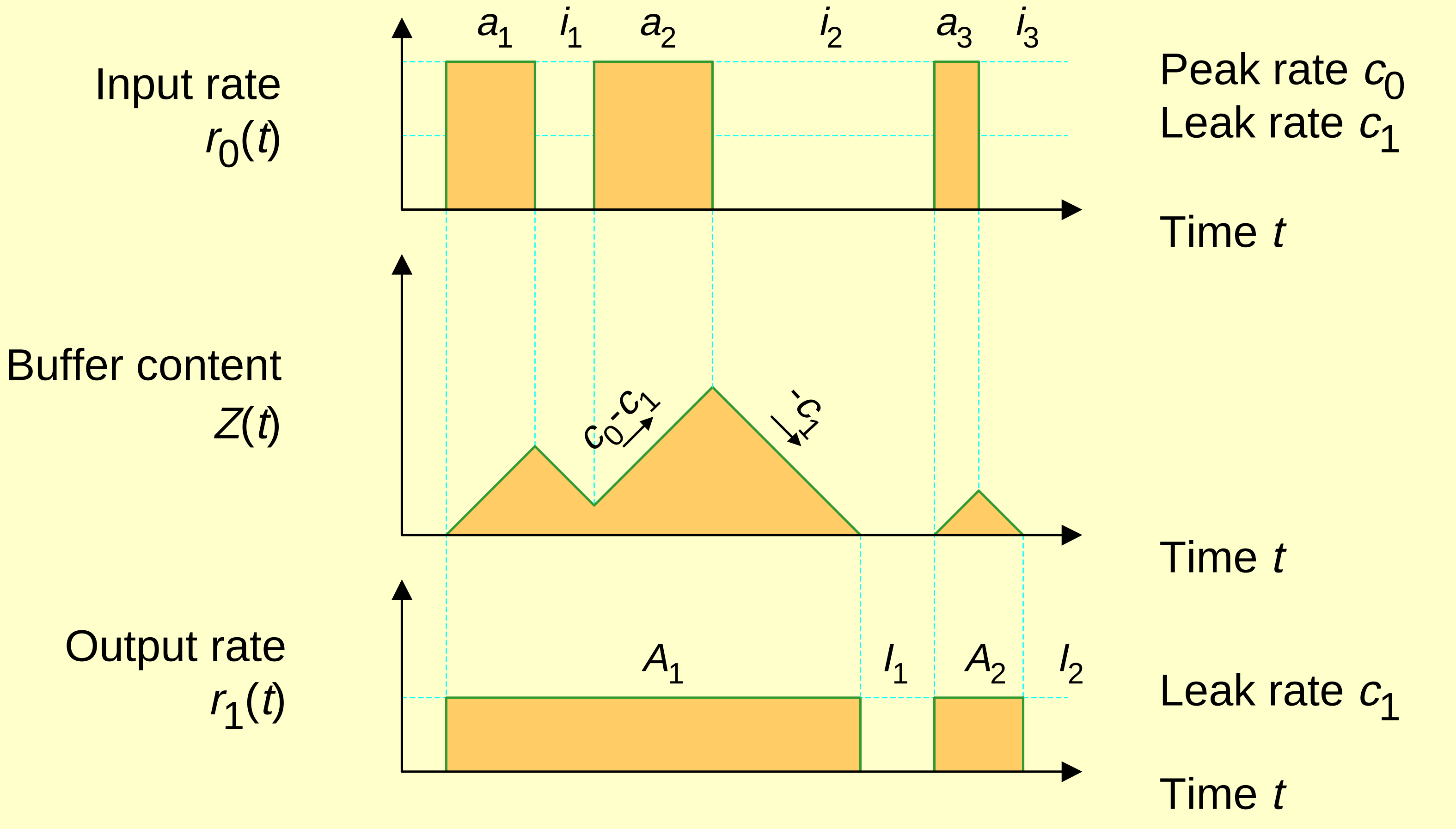
=> output looks like another on-off source (with rate c_1)

- idle periods ~ Exp(λ)
- active periods?

- Notation:

$$c = \frac{c_1}{c_0}$$

Evolution



Input rate
$r_0(t)$
a_1
i_1
a_2
i_2
a_3
i_3
Peak rate c_0
Leak rate c_1
Time t
Buffer content
$Z(t)$
c_0-c_1
-c_1
Time t
Output rate
$r_1(t)$
A_1
I_1
A_2
I_2
Leak rate c_1
Time t

Characterization of the output process

- Result:
  - active periods *A* on the output line ~
    busy periods *B* in an M/G/1 queue with
    arrival rate (1 - *c*)λ and service time distribution function *F*(*ct*)
- Proof:

$$A = \inf_{n \geq 1} \left\{ \sum_{m=1}^{n} \frac{a_m}{c} \,\middle|\, \sum_{m=1}^{n} \frac{a_m}{c} < \sum_{m=1}^{n} (a_m + i_m) \right\}$$

Since

$$\sum_{m=1}^{n} \frac{a_m}{c} < \sum_{m=1}^{n} (a_m + i_m) \Leftrightarrow \sum_{m=1}^{n} a_m \frac{1-c}{c} < \sum_{m=1}^{n} i_m$$

we get

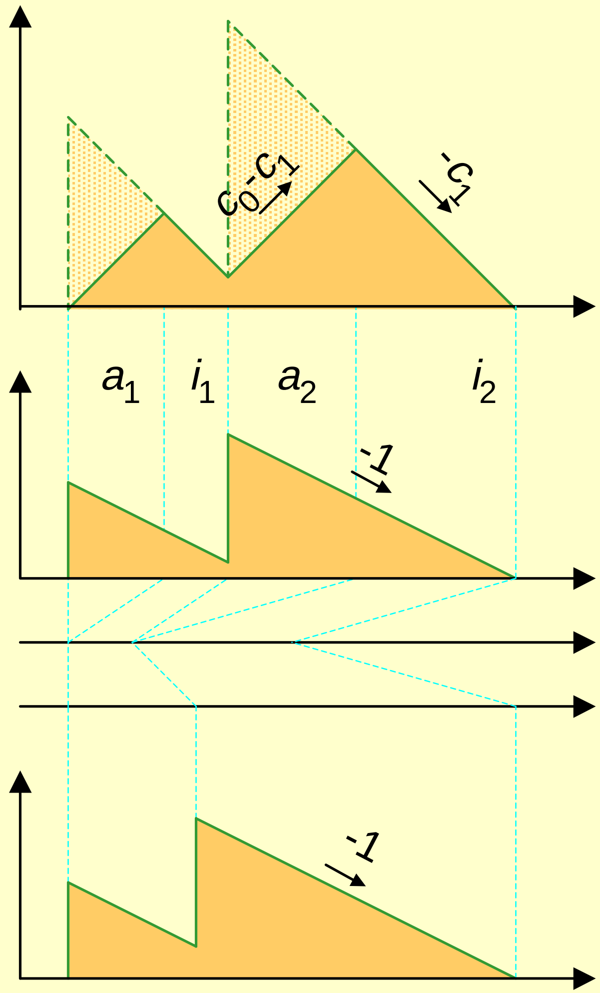
$$A = \inf_{n \geq 1} \left\{ \sum_{m=1}^{n} \frac{a_m}{c} \,\middle|\, \sum_{m=1}^{n} \frac{a_m}{c} < \sum_{m=1}^{n} \frac{i_m}{1-c} \right\}$$

Construction of the related M/G/1 workload process

Buffer content $Z(t)$

Workload $U(t)$

Workload $U'(t)$



Service times: a_m/c
Interarr. times: a_m+i_m

Service times: a_m/c
Interarr. times: $i_m/(1-c)$

Distribution of the buffer content

- Result:
  - tail distribution of the buffer content $Z \propto$ tail distribution of the workload V in an M/G/1 queue with arrival rate λ/c_1 and service time distribution function $F(z/(c_0-c_1))$:

$$P\{Z > z\} = \gamma \cdot P\{V > z\}$$

- Idea of the proof:
  - $Z_i(t)$ = buffer content during idle periods of the source
    - in the beginning of an idle period: jumps up $a_n(c_0 - c_1)$
    - during idle period: decreases with rate $-c_1$
  - $V(c_1 t) = Z_i(t)$

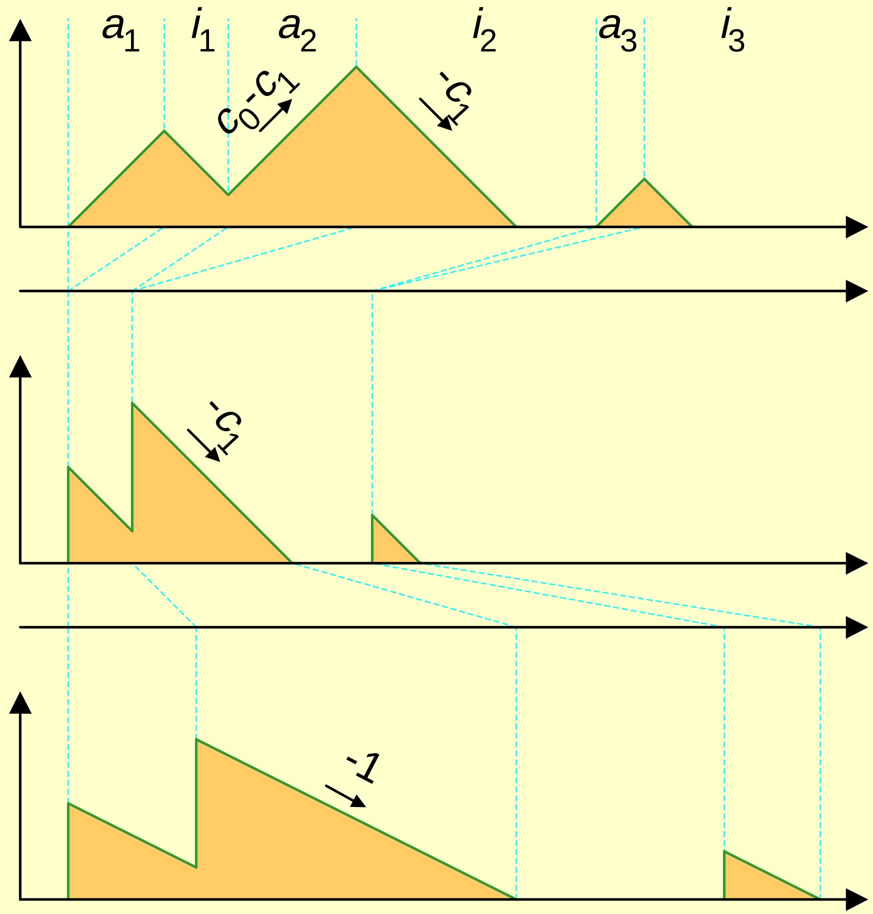
$$P\{Z > z\} \propto P\{Z_i > z\} = P\{V > z\}$$

Construction of the related M/G/1 workload process

Buffer content $Z(t)$

Buffer content $Z_i(t)$

Workload $V(t)$



Mean and variance of the buffer content

- Let

$$\beta_k = E[a^k] \qquad \rho = \frac{c_0}{c_1}\frac{\lambda\beta_1}{1+\lambda\beta_1}$$

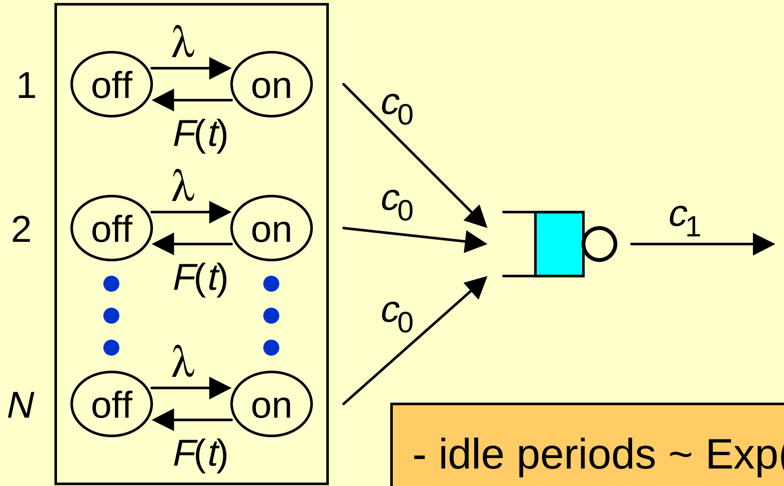
- Since E[Z^k] = $\gamma\cdot$ E[V^k], we have (whenever ρ < 1)

$$E[Z] = \frac{\beta_2}{2\beta_1}\frac{c_0 - c_1}{1+\lambda\beta_1}\frac{\rho}{1-\rho}$$

$$D^2[Z] = \frac{\beta_3}{3\beta_1}\frac{(c_0-c_1)^2}{1+\lambda\beta_1}\frac{\rho}{1-\rho} + 2\left(\frac{\beta_2}{2\beta_1}\right)^2\frac{(c_0-c_1)^3}{1+\lambda\beta_1}\left(\frac{\rho}{1-\rho}\right)^2\left(\frac{1}{c_0}-\frac{1}{2(c_0-c_1)(1+\lambda\beta_1)}\right)$$

Contents

Fluid queue driven by multiple homogeneous on-off sources



- Assumption:

$$c_0 \geq c_1$$

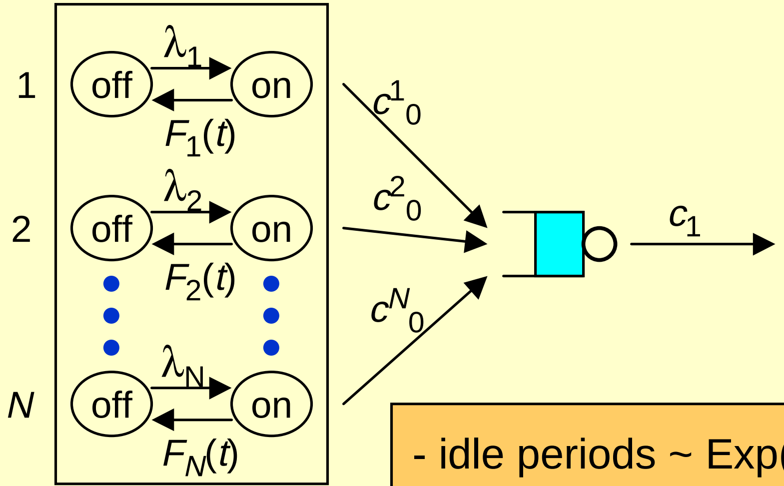
=> output looks like another on-off source (with rate c_1)

- Rubinovitch (1973): $c_0 = c_1$ =>
  - active periods on the output line ~ busy periods in an M/G/1 queue with arrival rate $(N - 1)\lambda$ and service time distribution function $F(t)$

- idle periods ~ Exp($N\lambda$)
- active periods?

- Boxma & Dumas (1998), Aalto (1998) [4]:
  - active periods on the output line ~ busy periods in an M/G/1 queue with arrival rate $(N - c)\lambda$ and service time distribution function $F(ct)$

Fluid queue driven by multiple heterogeneous on-off sources



- Assumption:

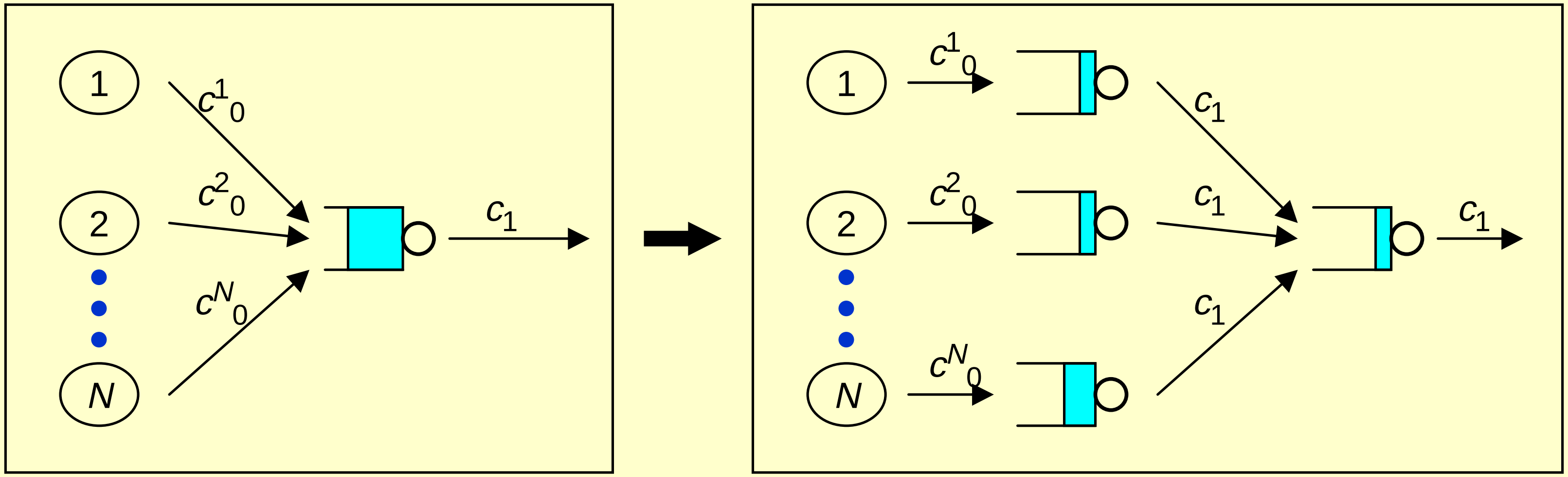
$$c_0^i \geq c_1 \ \text{ for all } i$$

=> output looks like another on-off source (with rate c_1)

- idle periods ~ Exp($\Sigma\lambda_i$)
- active periods?

- Kaspi & Rubinovitch (1975): $c^i_0 = c_1$ for all i =>
  - characterization of the active periods on the output line by Laplace transforms
- Boxma & Dumas (1998), Aalto (1998) [4]:
  - characterization of the active periods on the output line by Laplace transforms

Idea of the proof: insert intermediate buffers

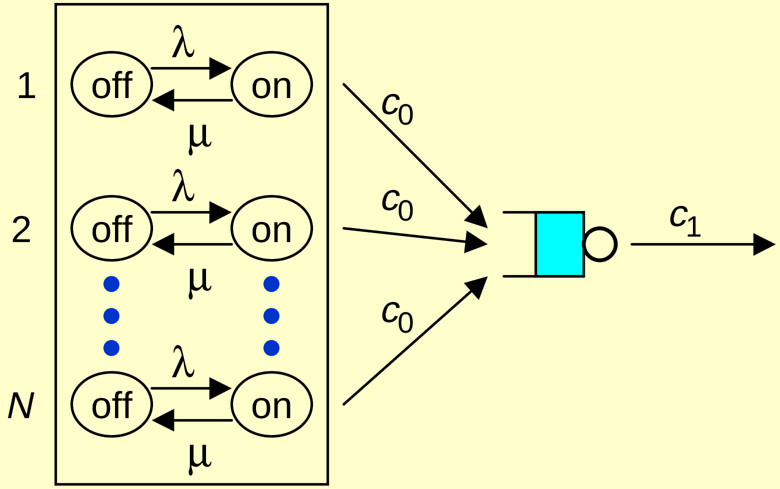


$$Z_1 = \tilde{Z}_{11} + \ldots + \tilde{Z}_{1N} + \tilde{Z}_2$$

=> outputs from the two systems are identical!

Contents

Fluid queue driven by multiple exponential on-off sources (1)



- **A-M-S model** first analysed by Anick, Mitra & Sondhi (1982)
- Assumption:

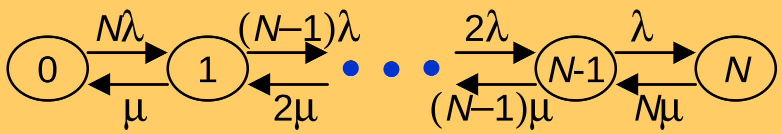
$$c_0 \neq c_1$$

- Input rate is modulated by a finite state Markov birth-death process $J(t)$: $r_0(t) = J(t)c_0$
- Aalto (1994) ITC-14:
  - characterization of the output rate process by constructing another Markov jump process, which modulates the output rate:

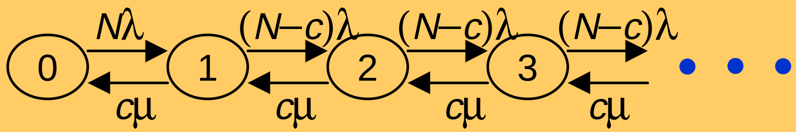
$$r_1(t) = d(\tilde{J}(t))$$

Fluid queue driven by multiple exponential on-off sources (2)

- Input rate is modulated by the following birth-death process $J(t)$:

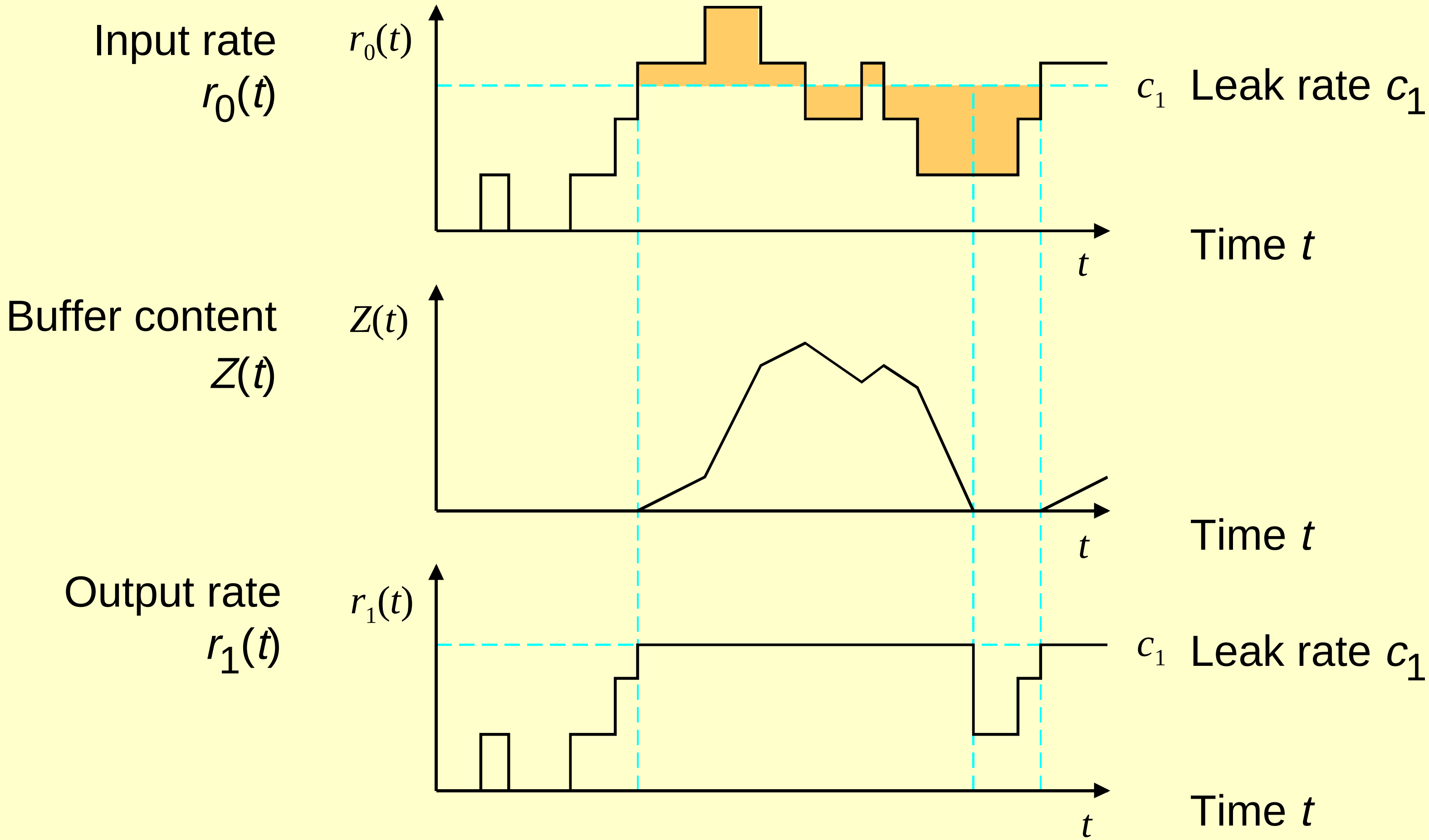


- In general, the ouput rate is modulated by a 3-dimensional Markov jump process (with an infinite state space)
- However, if $c_0 > c_1$, then the output looks like an on-off source and it is modulated by the following birth-death process:



  – note that, the active periods of output line ~ busy periods in an M/M/1 queue with parameters $(N - c)\lambda$ and $c\mu$

Idea of the proof: overloaded and underloaded intervals



Fluid queue driven by a Markov jump process (1)

- Input rate modulated by a (general) Markov jump process *J*(*t*)

$$r_0(t) = f_0(J(t))$$

- Assumption 1:

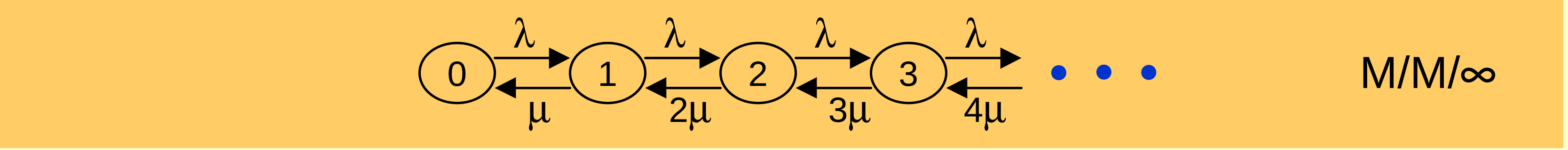
$$f_0(j) \neq c_1 \quad \text{for all } j$$

- Assumption 2:
  - visits to underloaded ($f_0(j) < c_1$) and overloaded ($f_0(j) > c_1$) states constitute an alternating renewal process
- Aalto (1998) [3]:
  - characterization of the output rate process by constructing another Markov jump process, which modulates the output rate:

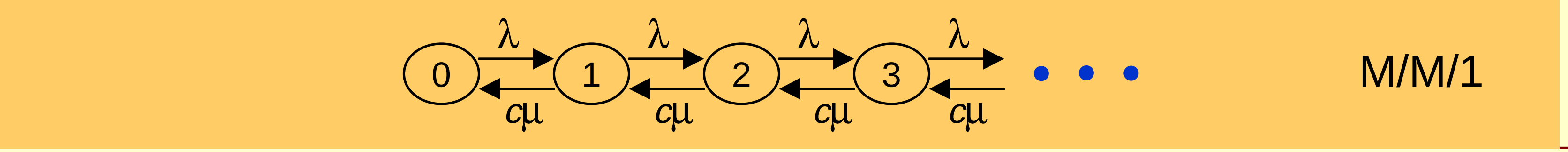
$$r_1(t) = f_1(\tilde{J}(t))$$

Fluid queue driven by a Markov jump process (2)

- In general, the ouput rate is modulated by a (2*d*+1)-dimensional Markov jump process (with an infinite state space), where *d* refers to the dimension of $J(t)$
- **Example** (Fluid queue driven by an M/M/∞ queue):
  - Input rate is modulated by the following birth-death process $J(t)$:



  - $r_0(t) = J(t)c_0$
  - If $c_0 > c_1$, then the output looks like an on-off source and it is modulated by the following birth-death process:

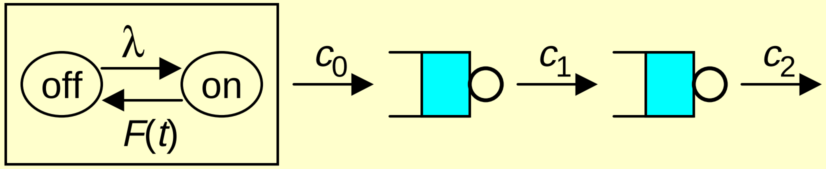


Contents

- Fluid queue
- Fluid queues driven by a single on-off source
- Fluid queues driven by multiple on-off sources
- Fluid queues driven by Markov jump processes
- Tandem fluid queues (joint work with Werner Scheinhardt)

Tandem fluid queue driven by a single on-off source

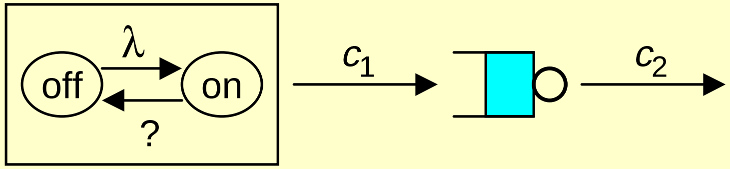
- On-off source with rate c_0
  - Idle periods ~ Exp(λ)
  - Active periods ~ $F(t)$



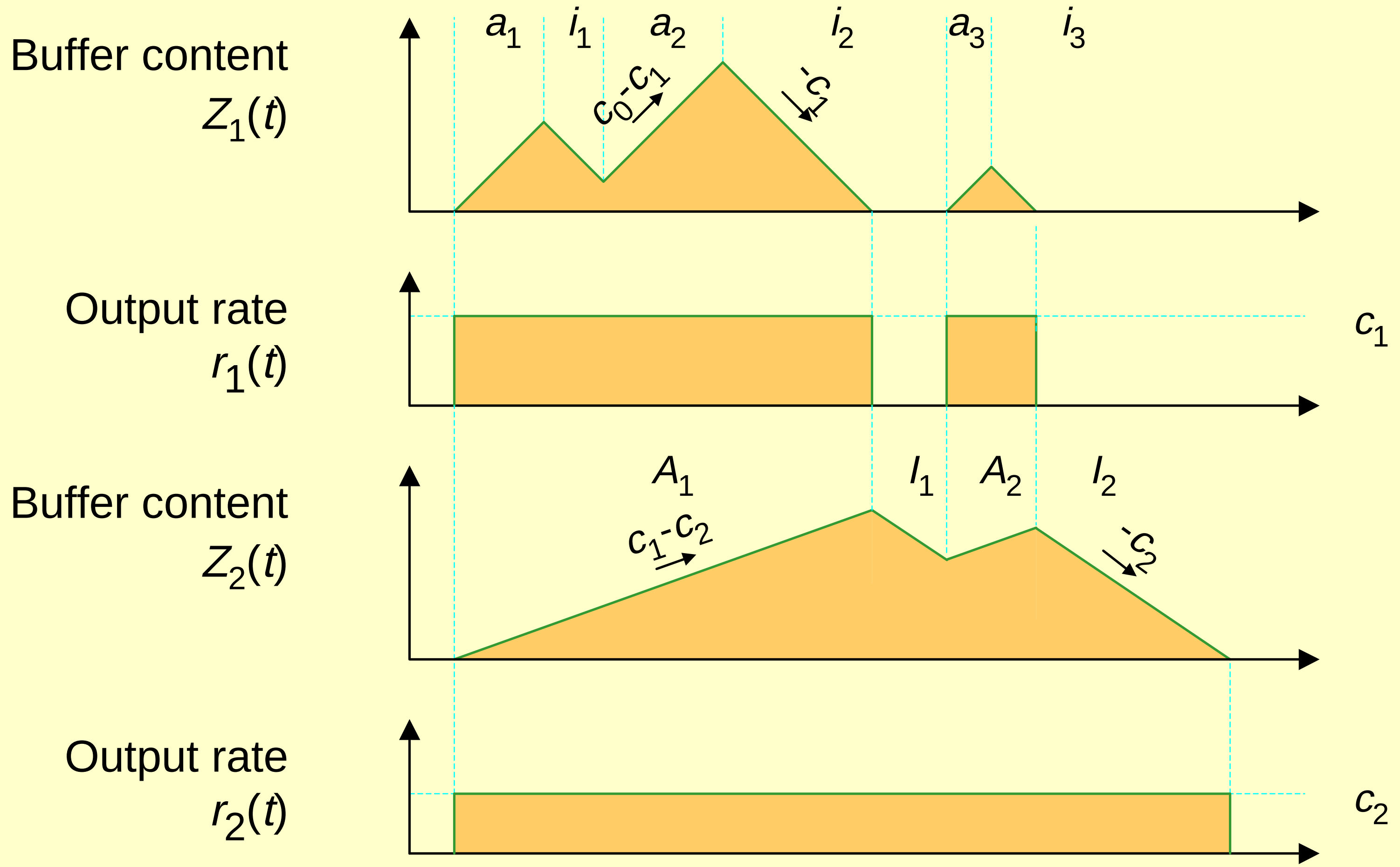
- $c_{i-1} \leq c_i$ => buffer i empty
- Assumption:

$$c_0 > c_1 > c_2$$

=> output from buffer i looks like another on-off source (with rate c_i)
- idle periods ~ Exp(λ)
- active periods?

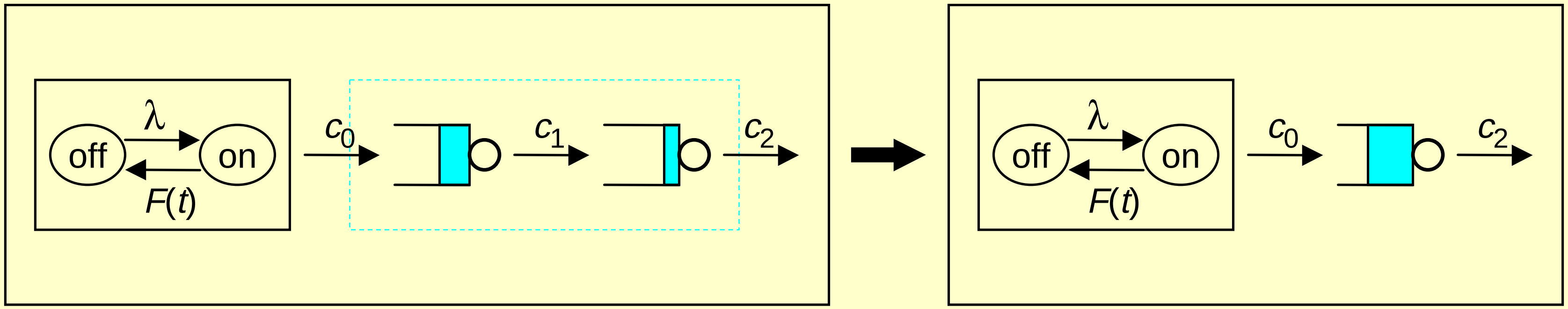


Evolution


Buffer content
Z1(t)
a1
i1
a2
i2
a3
i3
c0-c1
-c1
Output rate
r1(t)
c1
Buffer content
Z2(t)
A1
I1
A2
I2
c1-c2
-c2
Output rate
r2(t)
c2

Output from buffer *i*

- Result:
  - non-empty periods of buffer i ~
    busy periods in an M/G/1 queue with
    arrival rate $\lambda(c_0 - c_i)/c_0$ and service time distribution function $F(c_i t/c_0)$
- Idea of the proof: combine the two buffers



$$Z_1 \dashv Z_2 = \tilde{Z}_1$$

=> outputs from the two systems are identical!

Mean and variance of the content of buffer *i*

- Let

$$\beta_k = E[a^k] \qquad \tilde{\beta}_k = E[(ac_0)^k] \qquad \rho_i = \frac{c_0}{c_i}\frac{\lambda\beta_1}{1+\lambda\beta_1} \qquad \alpha_i = \frac{1-\rho_i}{\rho_i}$$

- Result: If $\rho_i < 1$, then

$$E[Z_i] = \frac{\tilde{\beta}_2}{2\tilde{\beta}_1}\left(\frac{1}{1+\lambda\beta_1}\right)^2 \frac{\alpha_{i-1}-\alpha_i}{\alpha_{i-1}\alpha_i}$$

$$D^2[Z_i] = \frac{\tilde{\beta}_3}{3\tilde{\beta}_1}\left(\frac{1}{1+\lambda\beta_1}\right)^3 \frac{(\alpha_{i-1}-\alpha_i)^2}{\alpha_{i-1}^2\alpha_i} + \left(\frac{\tilde{\beta}_2}{2\tilde{\beta}_1}\right)^2\left(\frac{1}{1+\lambda\beta_1}\right)^4 \frac{(\alpha_{i-1}-\alpha_i)^2}{\alpha_{i-1}^3\alpha_i^2}(\alpha_{i-1}+2\alpha_i-4\lambda\beta_1\alpha_{i-1}\alpha_i)$$

Correlation between the buffer contents

- Let

$$B = \frac{4\beta_3\beta_1}{3\beta_2^2}(1+\lambda\beta_1) - 4\lambda\beta_1$$

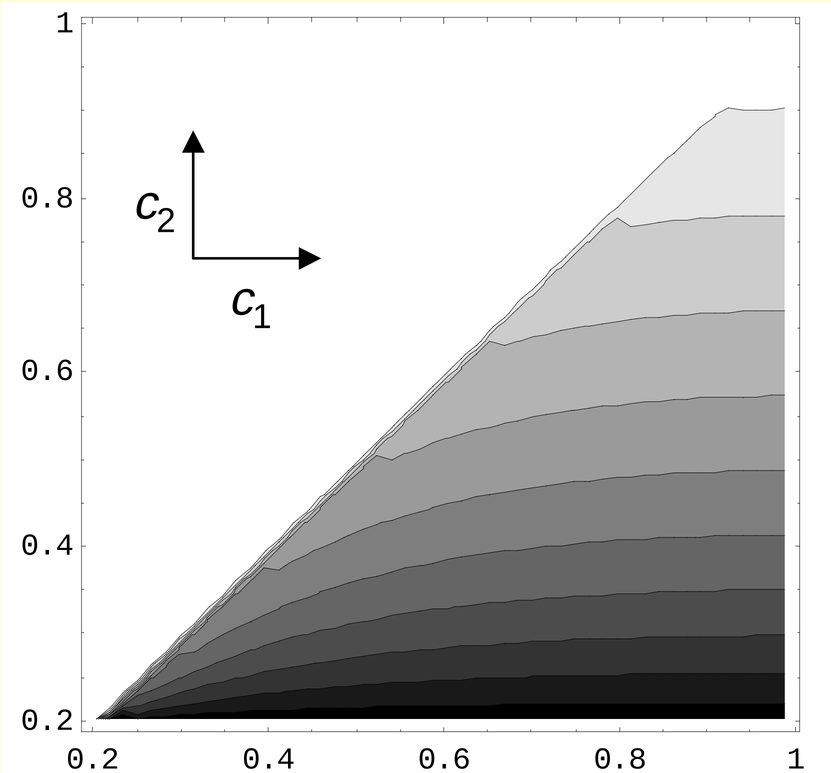
- Result: If $\rho_2 < 1$, then

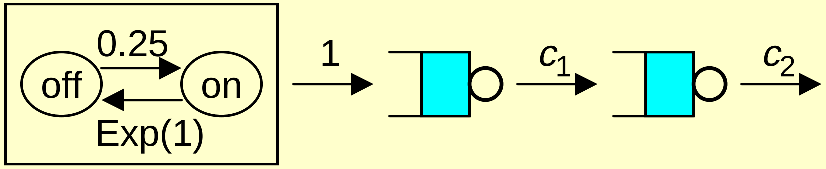
$$\mathrm{Corr}[Z_1, Z_2] = \frac{B\frac{(\alpha_0+\alpha_1)\alpha_1\alpha_2}{2} + \frac{(\alpha_0^2+\alpha_0\alpha_1+\alpha_1^2)\alpha_2}{\alpha_0}}{\sqrt{B\alpha_0^2\alpha_1+\alpha_0(\alpha_0+2\alpha_1)}\sqrt{B\alpha_1^2\alpha_2+\alpha_1(\alpha_1+2\alpha_2)}}$$

- Idea of the proof:

$$2\,\mathrm{Cov}[Z_1, Z_2] = D^2[Z_1+Z_2] - D^2[Z_1] - D^2[Z_2]$$

Example (1)

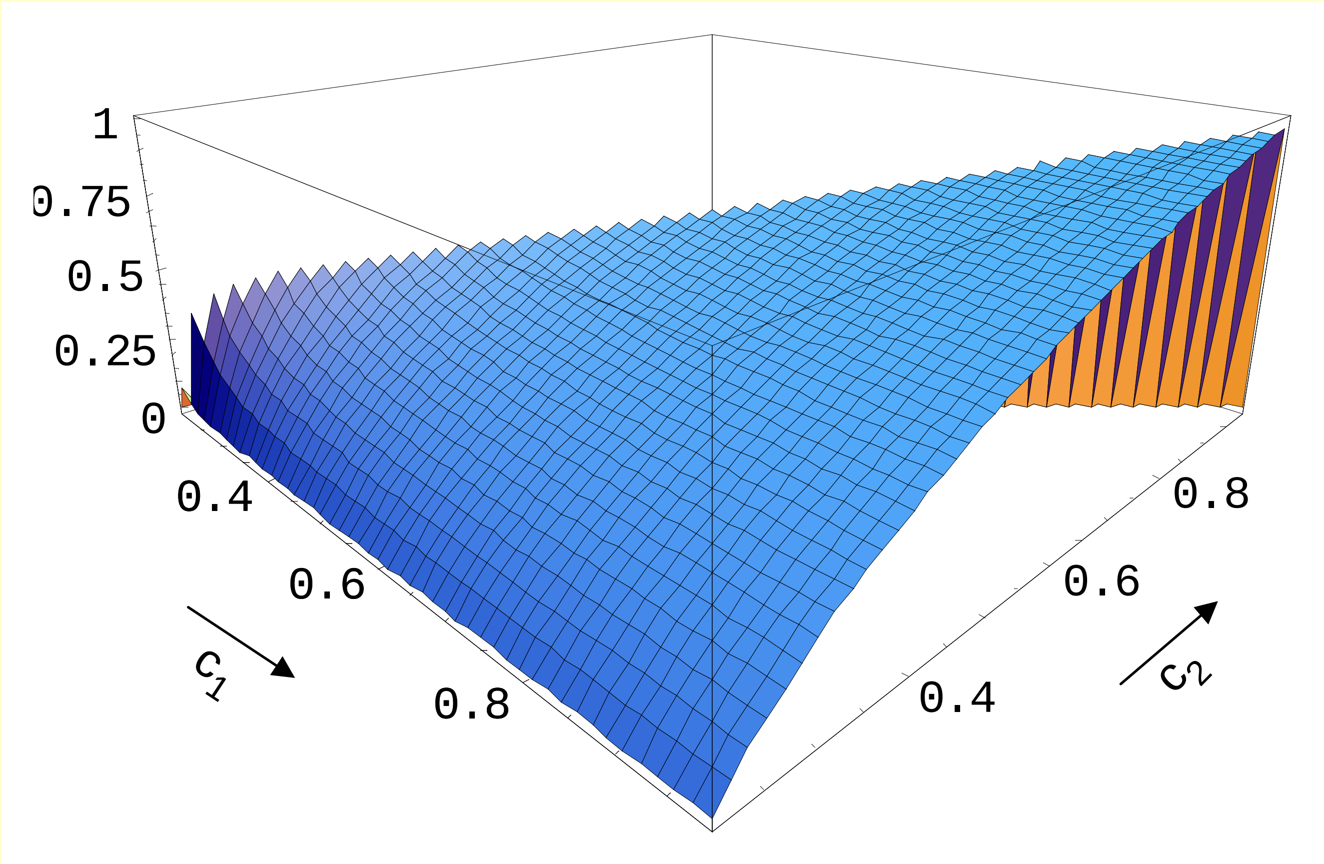




Stability condition:

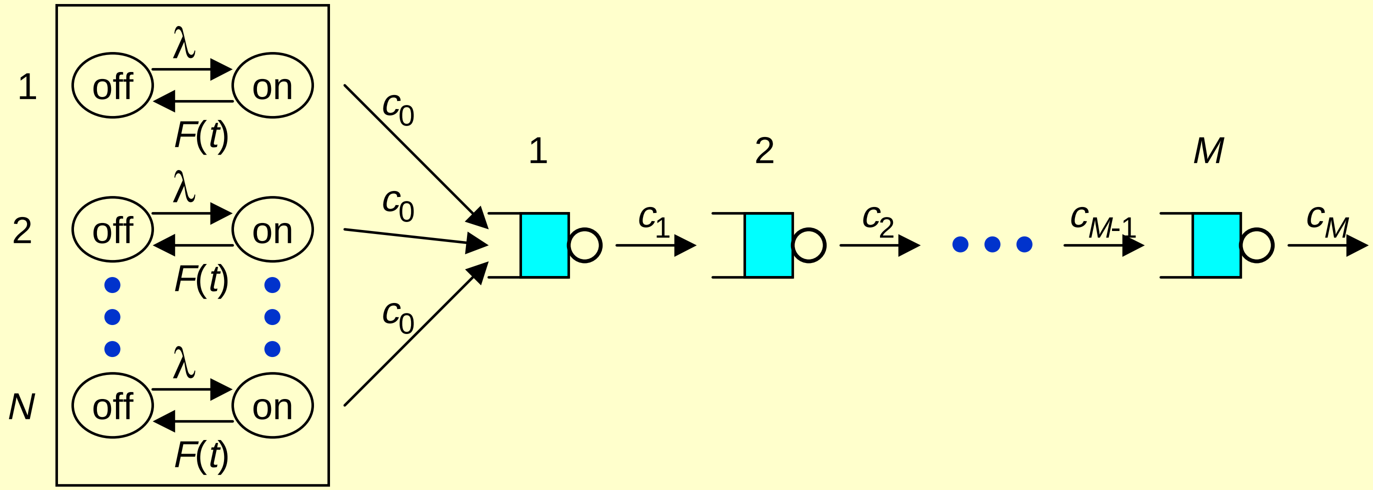
$$1 = c_0 > c_1 > c_2 > 0.2$$

Example (2)



Generalization

- Tandem fluid queue driven by multiple on-off sources



- Assumption:

$$c_0 > c_1 > c_2 > \ldots > c_M$$

=> output from buffer *i* looks like another on-off source

LOPPU