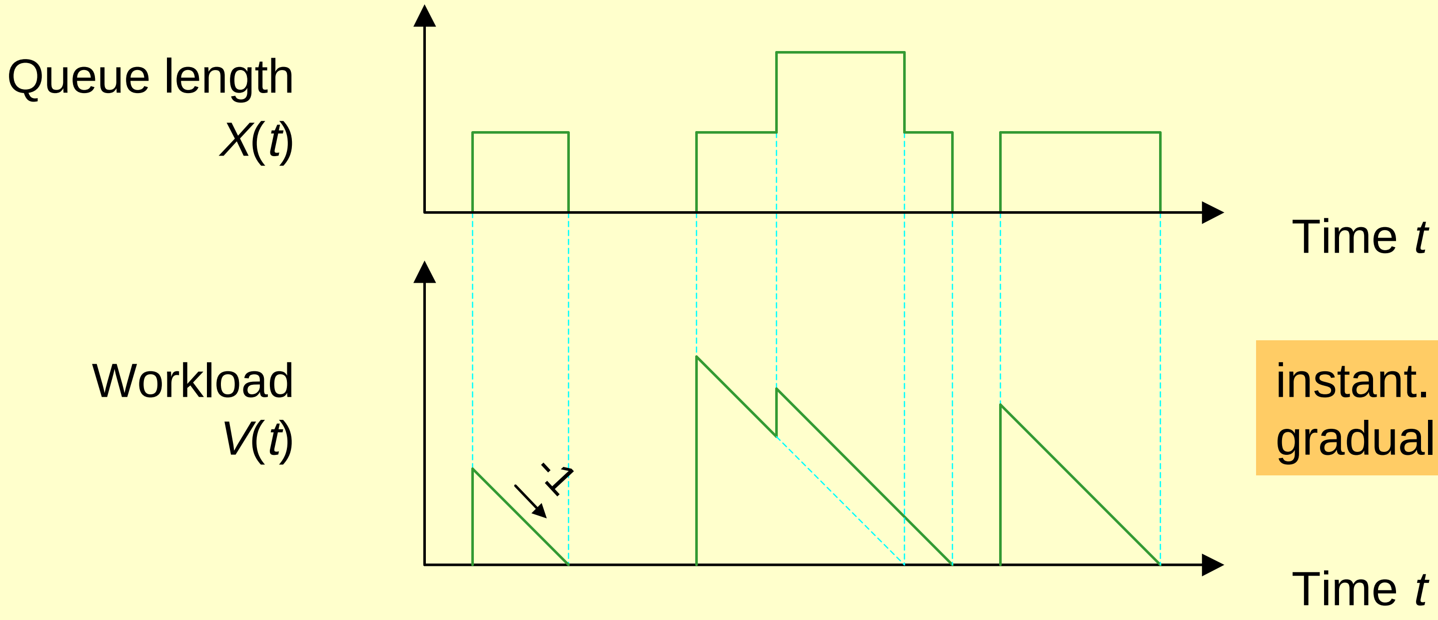
Characterization of the output process for some fluid queues

Samuli Aalto
EURANDOM

Background: ordinary single server queue



instant. input
gradual output

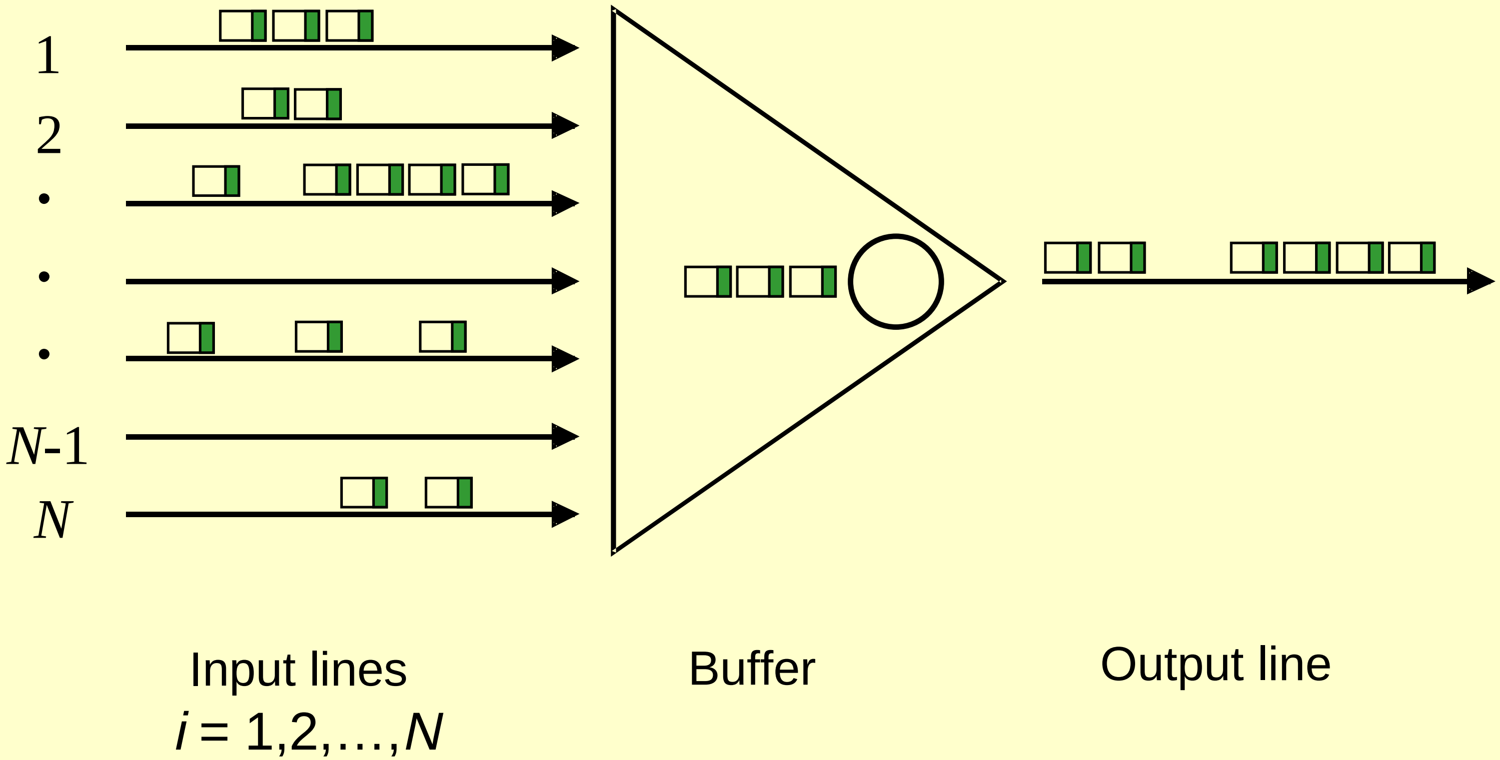
- Busy period *B*:

$$B = \inf_{n\geq 1} \left\{ \sum_{m=1}^{n} x_m \,\middle|\, \sum_{m=1}^{n} x_m < \sum_{m=1}^{n} t_m \right\}$$

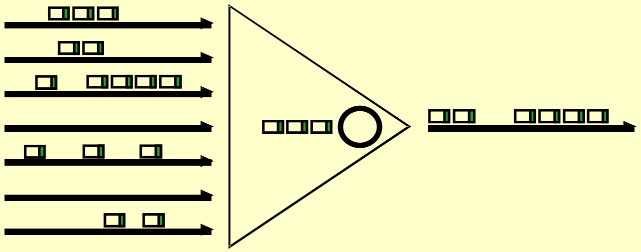
Contents

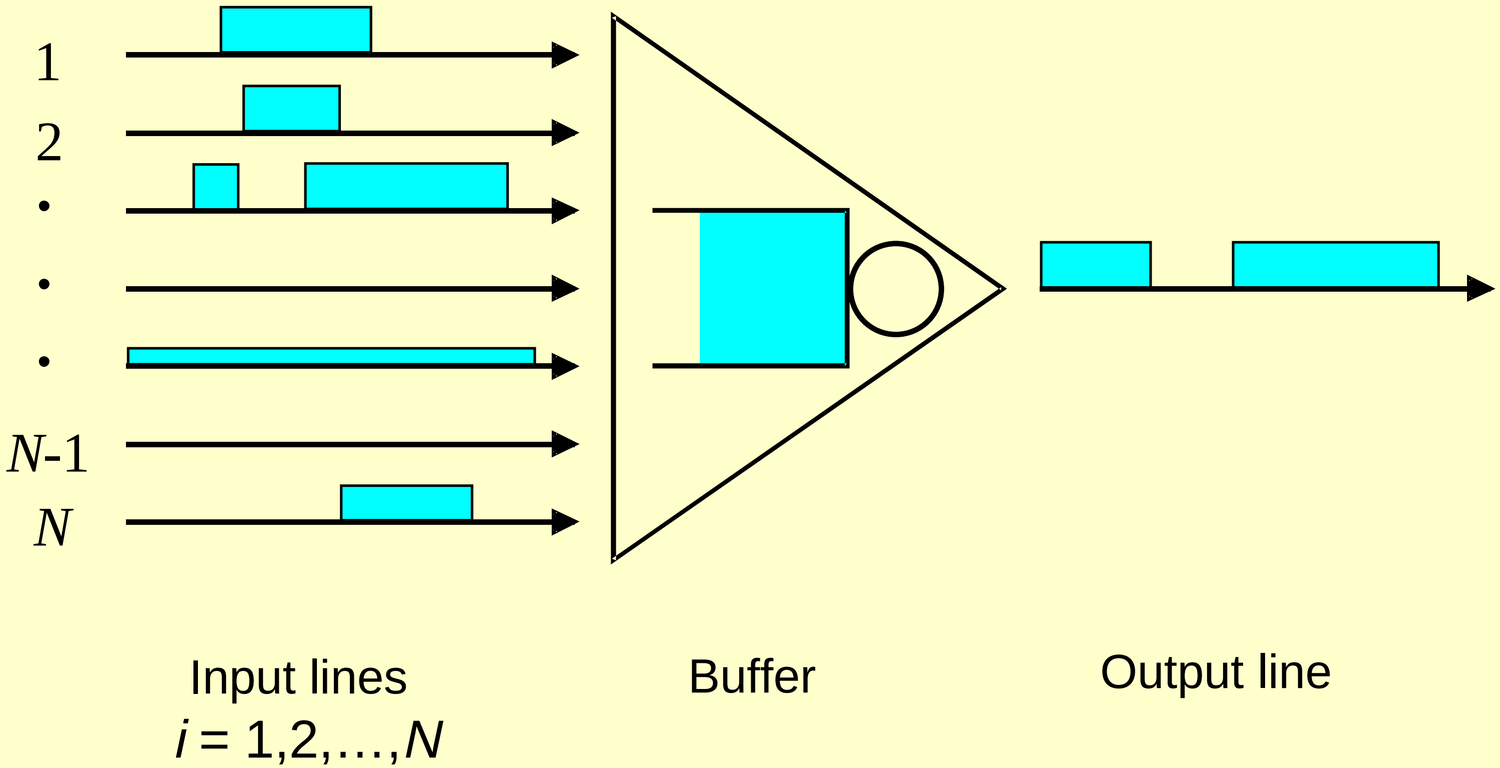
- Fluid queue
- Fluid queues fed by a single on-off source
- Fluid queues fed by multiple on-off sources
- Fluid queues driven by Markov jump processes
- Tandem fluid queues

Statistical multiplexer ...

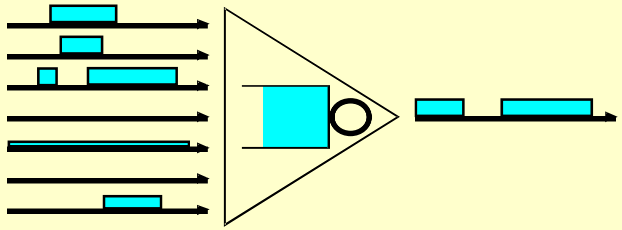


… as seen at the burst level





Fluid queue = fluid flow storage model



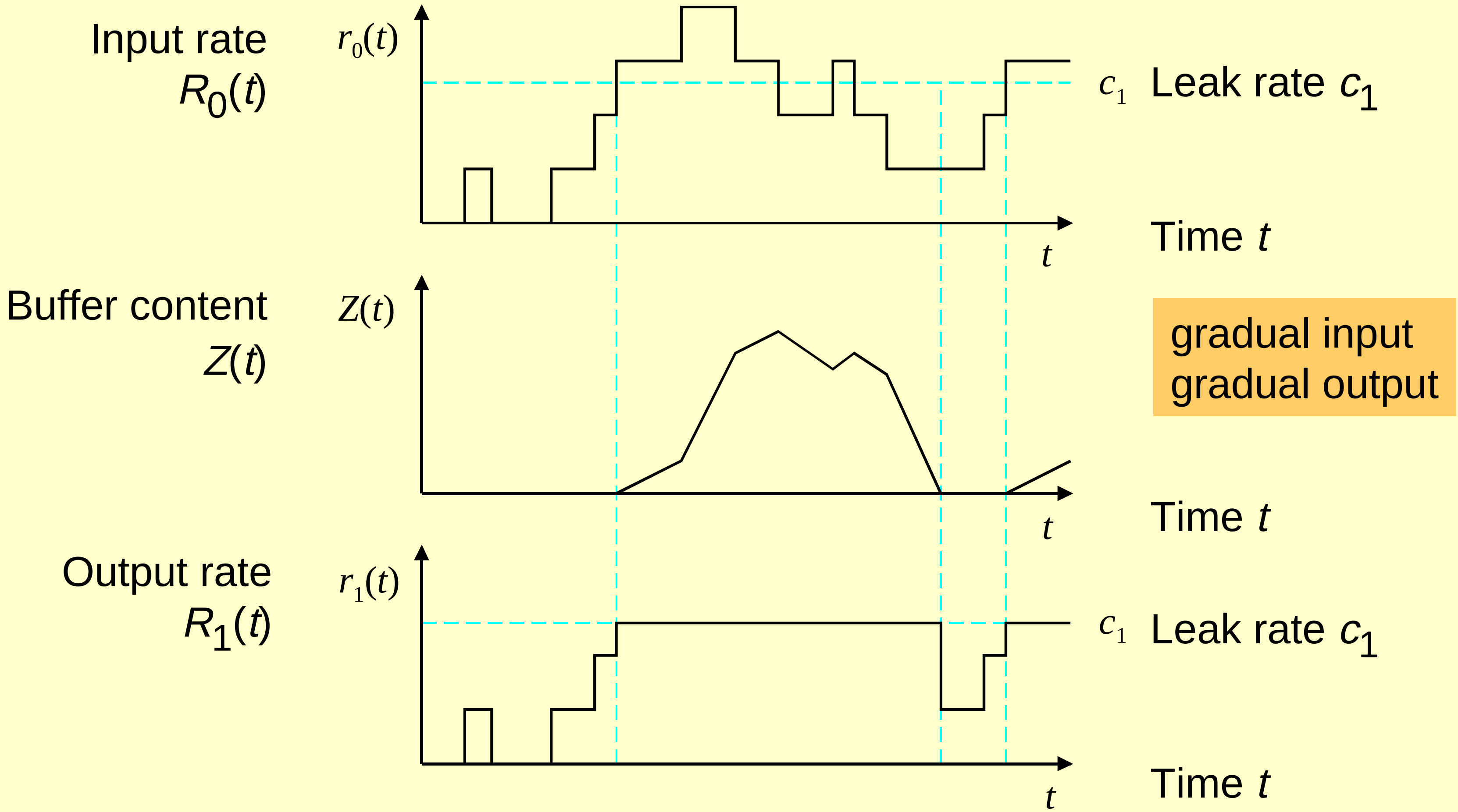
- Input rate $R_0(t)$
  - varying randomly
  - gradual input!
- Buffer size
  - we assume: infinite
- Leak rate c_1
  - max output rate
  - gradual output!

- Buffer content process $Z(t)$
- Output rate process $R_1(t)$

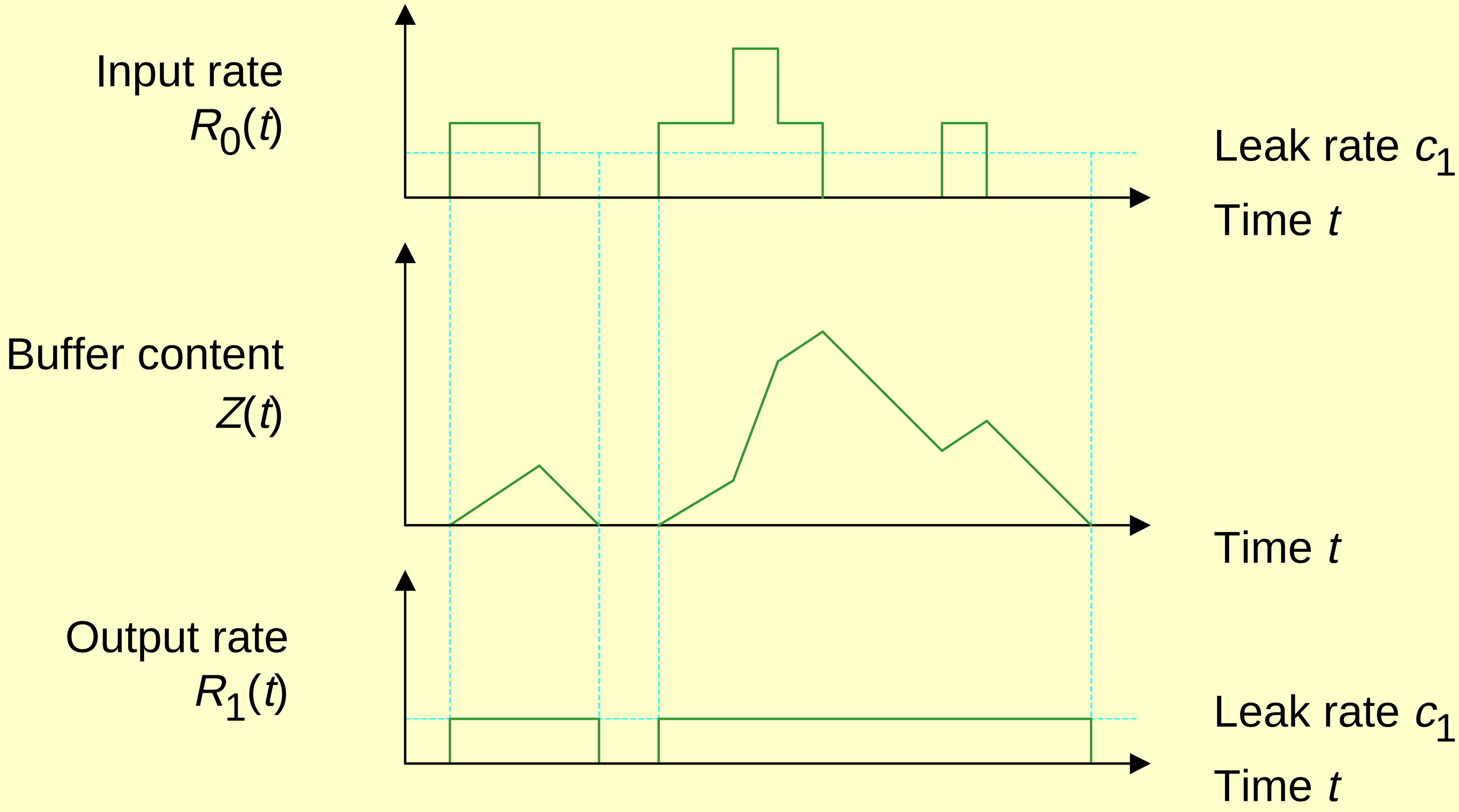
$$Z(t) = Z(0) + \int_0^t R_0(u)du - \int_0^t R_1(u)du$$

$$R_1(t) = \begin{cases} \min\{R_0(t), c_1\}, & \text{if } Z(t) = 0 \\ c_1 & \text{, if } Z(t) > 0 \end{cases}$$

Evolution

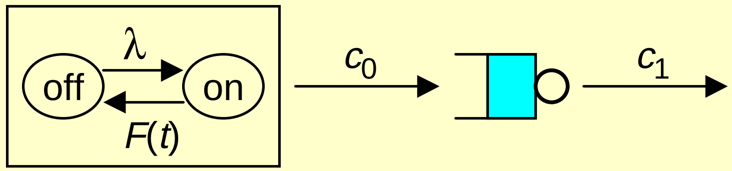

Input rate
$R_0(t)$
$r_0(t)$
c_1
Leak rate c_1
t
Time t
Buffer content
$Z(t)$
$Z(t)$
gradual input
gradual output
t
Time t
Output rate
$R_1(t)$
$r_1(t)$
c_1
Leak rate c_1
t
Time t

Evolution assumed that $R_0(t) > c_1$ whenever $R_0(t) > 0$



Contents

Fluid queue fed by a single on-off source



- On-off source with rate c_0
  - Silent periods ~ Exp(λ)
  - Active periods ~ $F(t)$
- $c_0 \leq c_1$ => input = output
- Assumption:

$$c_0 > c_1$$

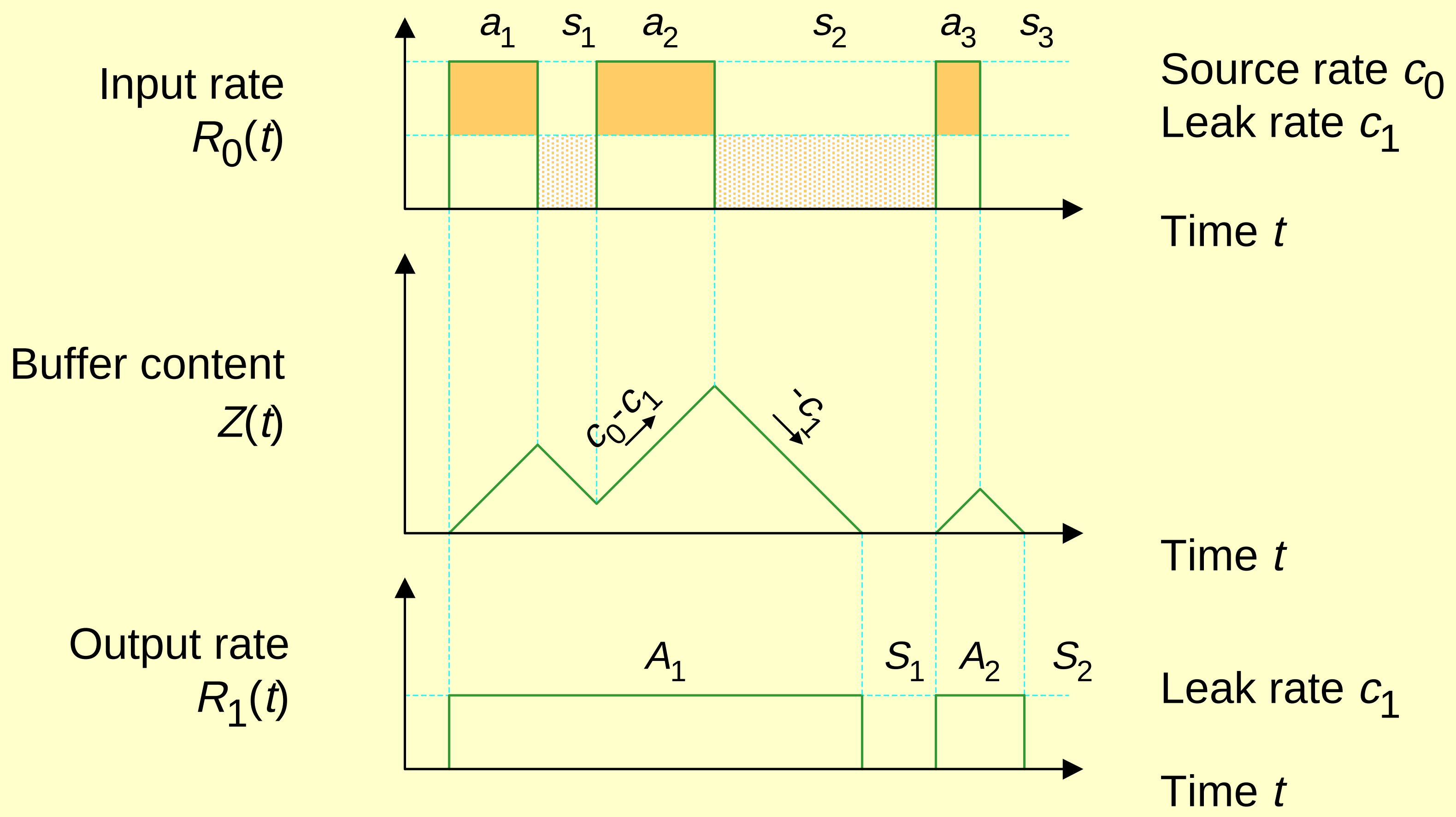
=> output looks like another on-off source (with rate c_1)

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

- Notation:

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

Evolution



Input rate
$R_0(t)$
a_1 s_1 a_2 s_2 a_3 s_3
Source 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 S_1 A_2 S_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 $(1 - c)\lambda$ and service time distribution function $F(ct)$
- Proof:

$$A = \inf_{n \geq 1} \left\{ \sum_{m=1}^{n} \frac{c_0 a_m}{c_1} \,\middle|\, \sum_{m=1}^{n} (c_0 - c_1) a_m < \sum_{m=1}^{n} c_1 s_m \right\}$$

Multiply by $c_0/(c_1(c_0 - c_1))$:

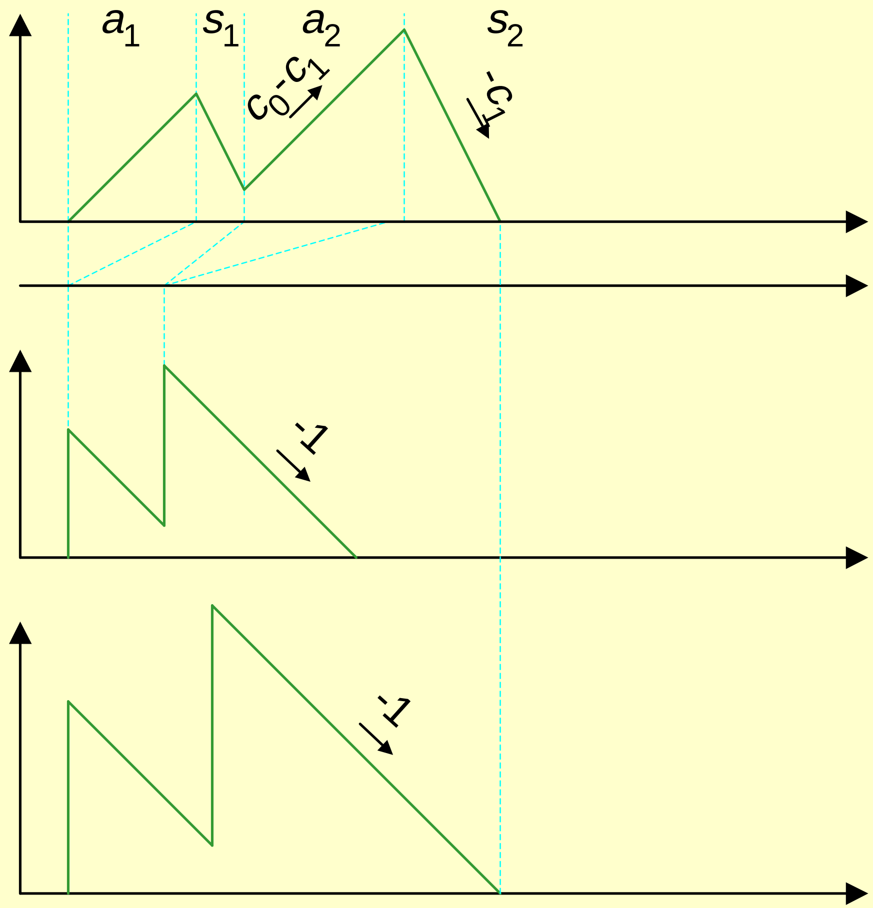
$$A = \inf_{n \geq 1} \left\{ \sum_{m=1}^{n} \frac{a_m}{c} \,\middle|\, \sum_{m=1}^{n} \frac{a_m}{c} < \sum_{m=1}^{n} \frac{s_m}{1-c} \right\}$$

Construction of the related M/G/1 workload process

Buffer content
$Z(t)$

Workload
$V(t)$

Workload
$V'(t)$



Stationary distribution of the buffer content

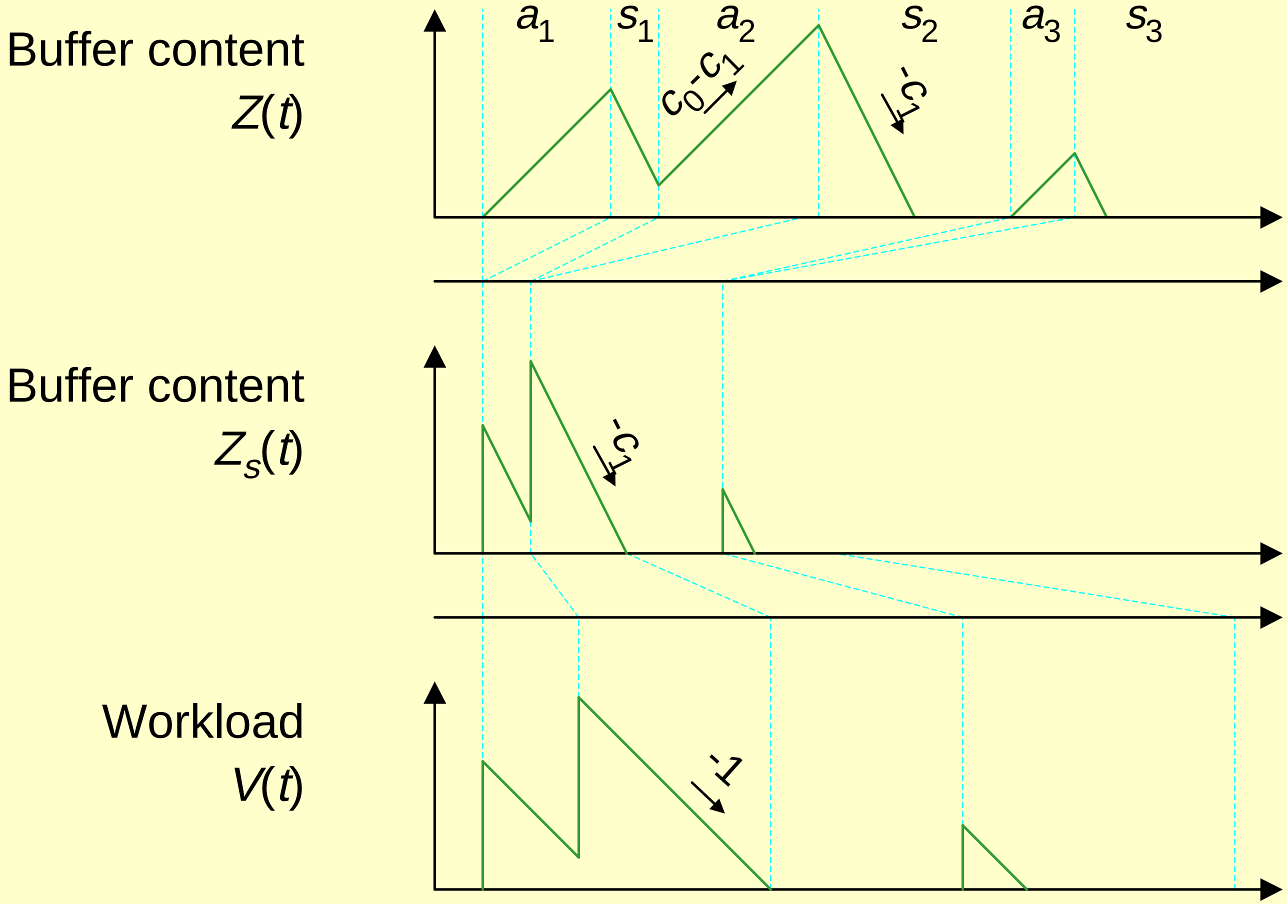
- Result:
  - tail distribution of the buffer content $Z \propto$ tail distribution of the workload V in an M/G/1 queue with arrival rate λ/c_1 and service time distribution function $F(z/(c_0-c_1))$:

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

- Idea of the proof:
  - $Z_s(t)$ = buffer content during silent periods of the source
    - in the beginning of silent period s_n : jumps up $a_n(c_0 - c_1)$
    - during silent period s_n: decreases linearly at rate $-c_1$
  - $V(t) = Z_s(t/c_1)$

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

Construction of the related M/G/1 workload process



First moments of the buffer content

- Let

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

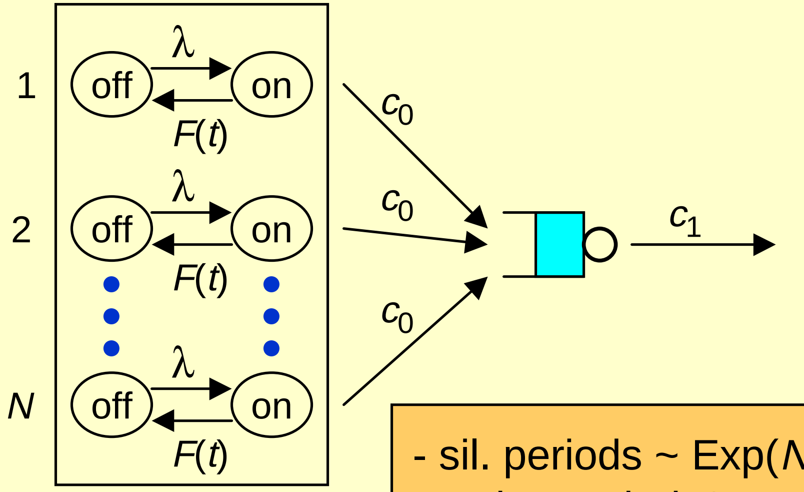
- Since E[Z^k] = $\gamma\cdot$E[V^k], we have (whenever $\rho < 1$)

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

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

Contents

Fluid queue fed by multiple homogeneous on-off sources



- Assumption:

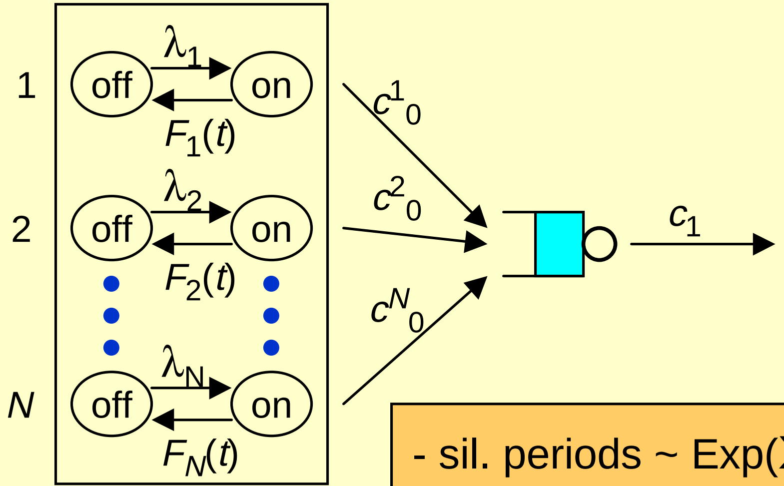
$$c_0 \geq c_1$$

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

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

- 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 $F(t)$
- 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 fed by multiple heterogeneous on-off sources



- Assumption:

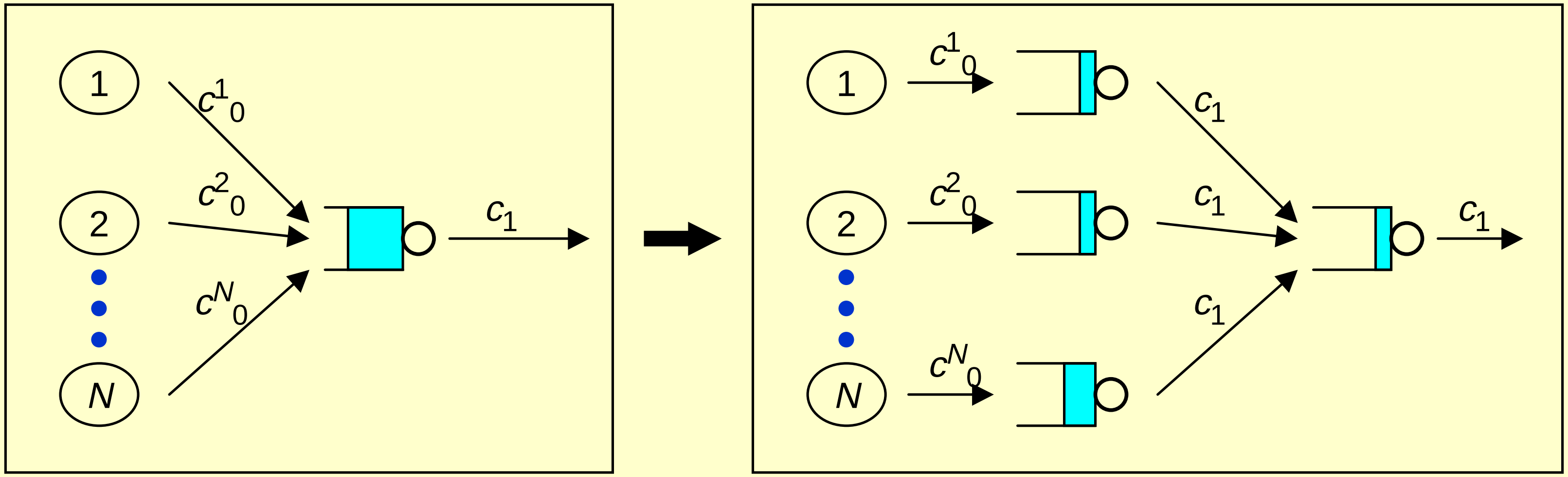
$$c_0^i \geq c_1 \quad \text{for all } i$$

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

- sil. 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 giving an implicit formula for the Laplace-Steltjes transform
- Boxma & Dumas (1998), Aalto (1998) [4]:
  - similar characterization

Idea of the proof: insert intermediate buffers



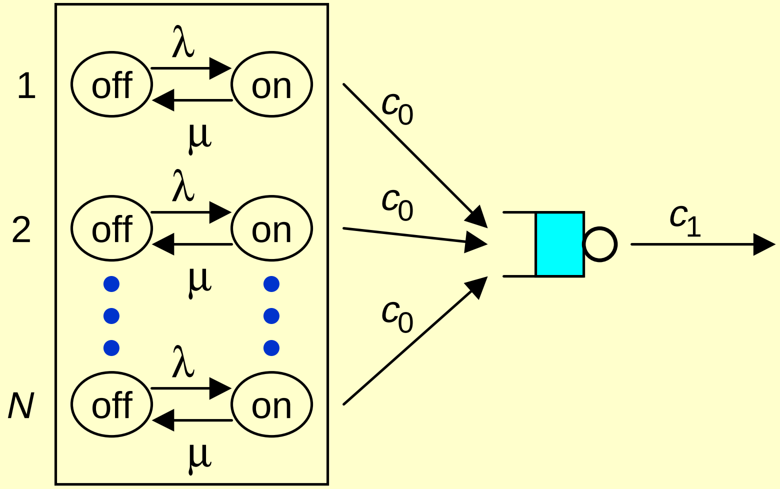
$$\forall t : Z_1(t) = \tilde{Z}_{11}(t) + \ldots + \tilde{Z}_{1N}(t) + \tilde{Z}_2(t)$$

=> outputs from the two systems are identical!

Contents

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

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



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

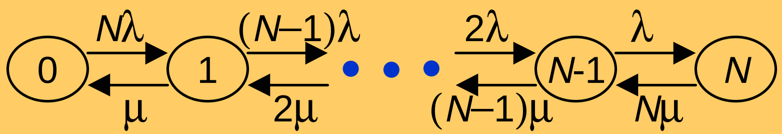
$$c_0 \neq c_1$$

- 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:

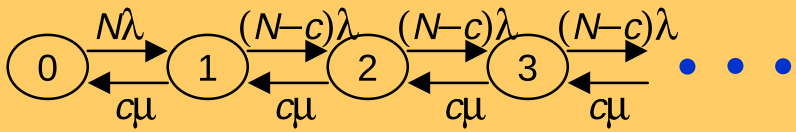
$$R_1(t) = f(\tilde{J}(t))$$

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

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

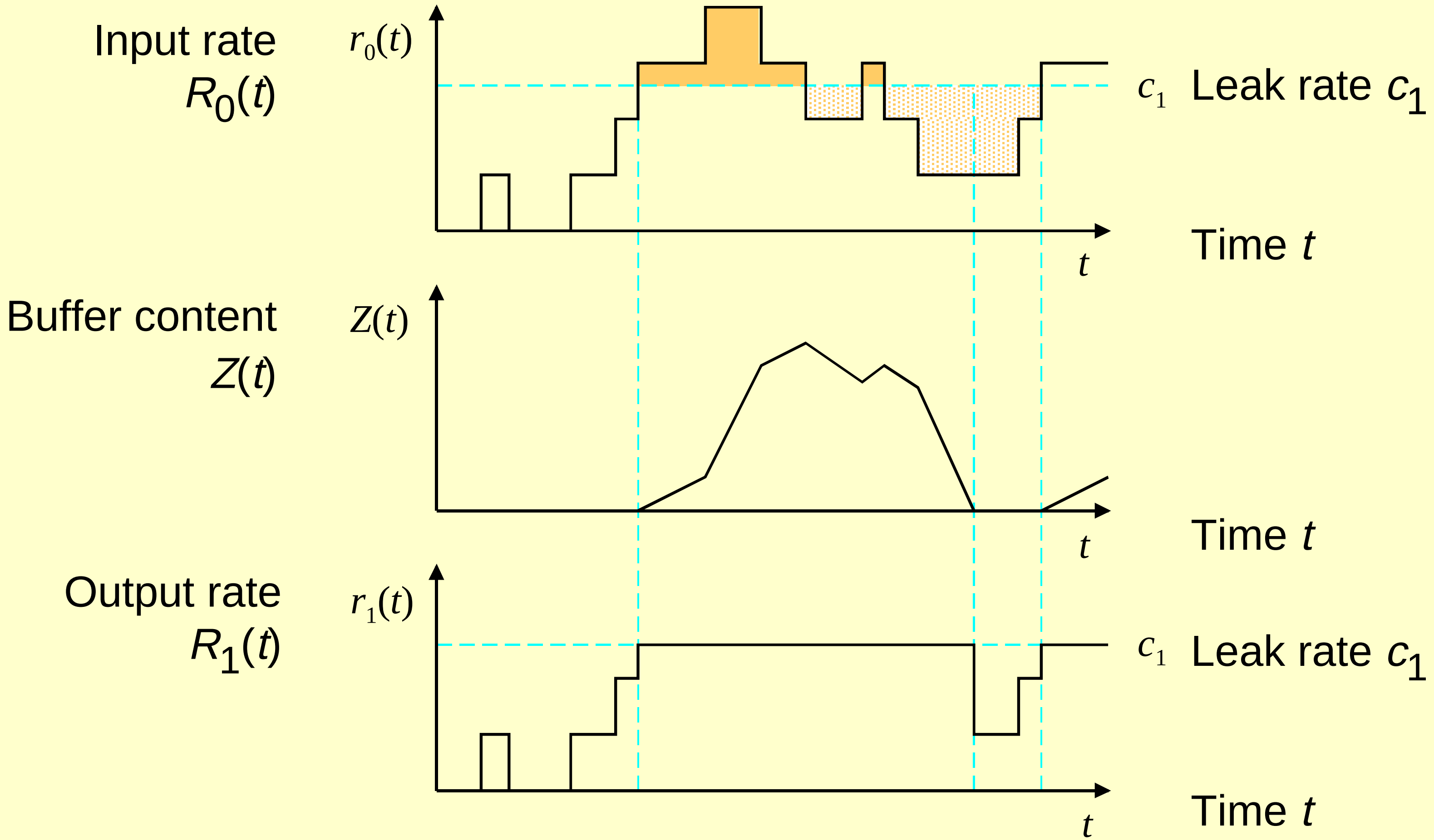


- In general, the ouput rate is modulated by a 3-dimensional Markov jump process (with an infinite state space)
- However, if $c_0 > c_1$, then the output looks like an on-off source and it is modulated by the following birth-death process:



  – 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:

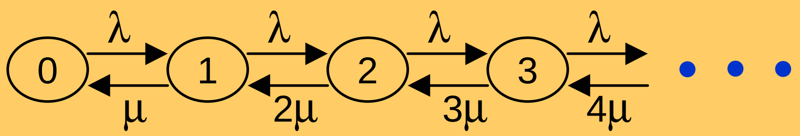
$$f_0(j) \neq c_1 \quad \text{for all } j$$

- 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:

$$R_1(t) = f_1(\tilde{J}(t))$$

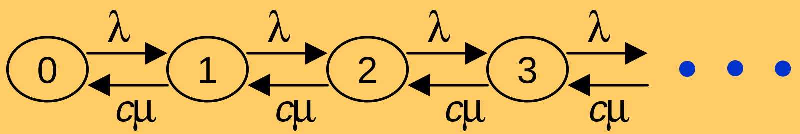
Fluid queue driven by a Markov jump process (2)

- In general, the ouput rate is modulated by a ($2d$+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 $J(t)$:



M/M/∞

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

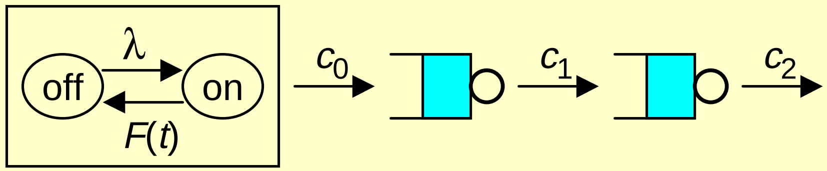


M/M/1

Contents

Tandem fluid queue fed by a single on-off source

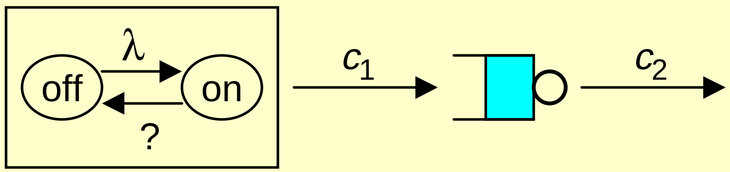
- On-off source with rate c_0
  - silent periods ~ Exp(λ)
  - active periods ~ $F(t)$
- $c_{i\text{-}1} \leq c_i$ => buffer i empty
- Assumption:



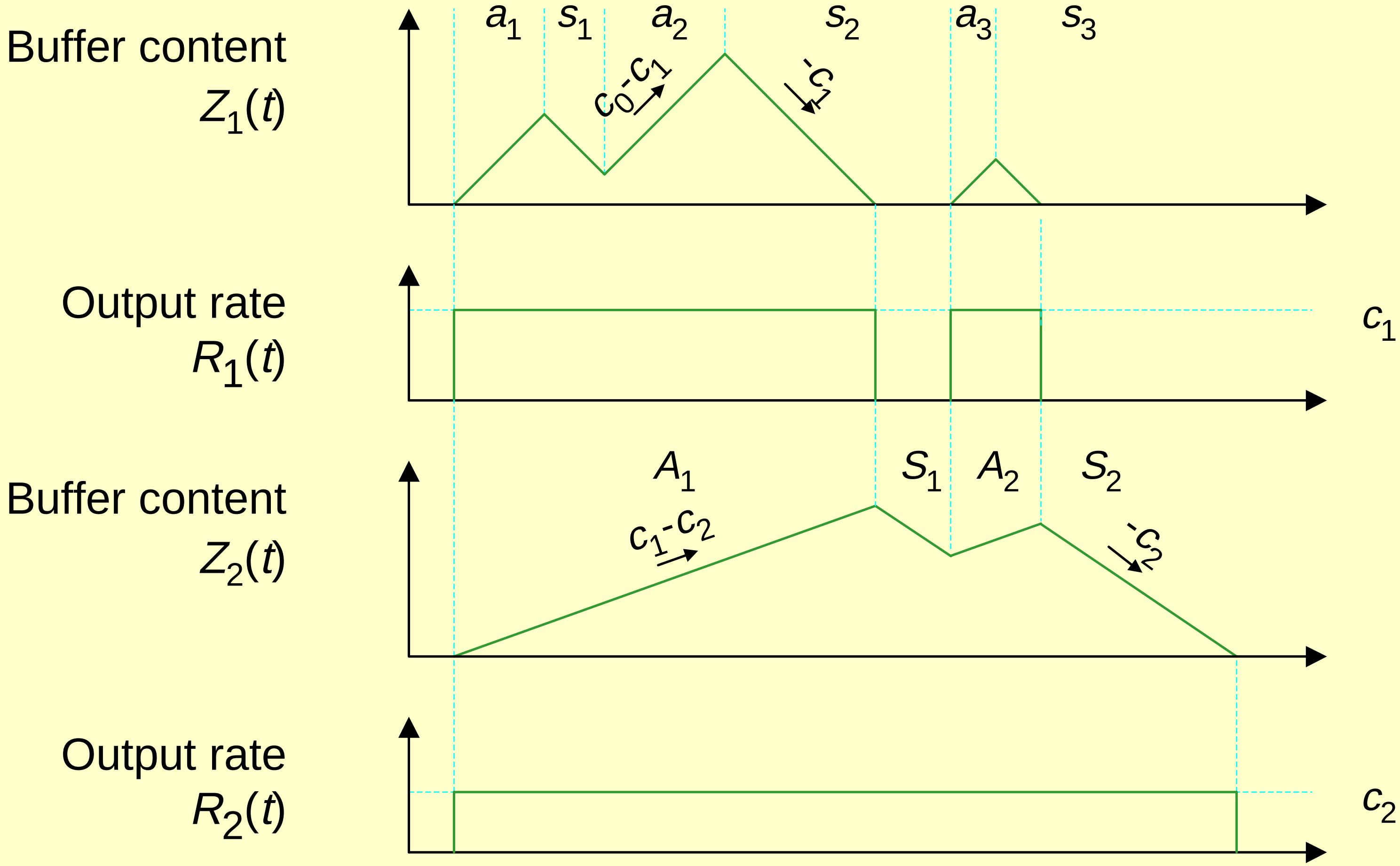
$$c_0 > c_1 > c_2$$

=> output from buffer i looks like another on-off source (with rate c_i)

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

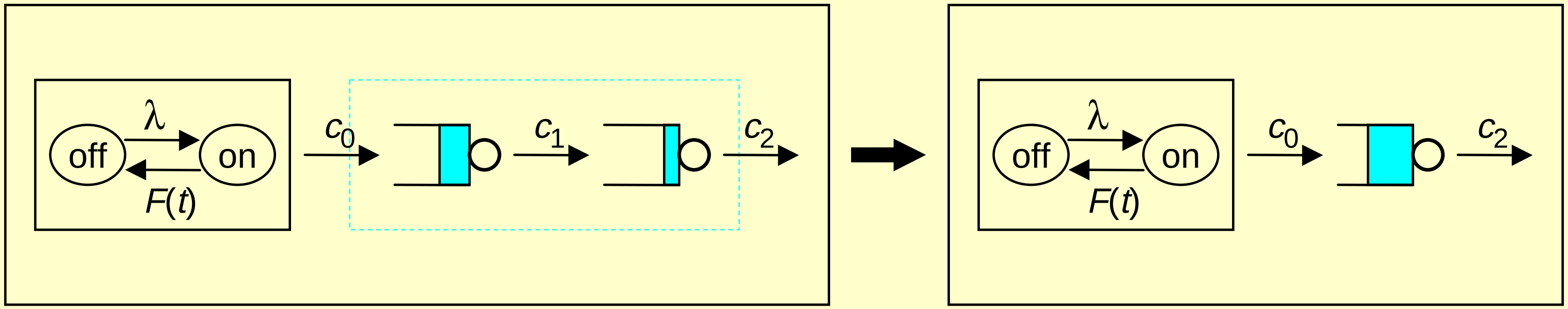


Evolution


Buffer content
$Z_1(t)$
a_1 s_1 a_2 s_2 a_3 s_3
c_0-c_1
$-c_1$
Output rate
$R_1(t)$
c_1
Buffer content
$Z_2(t)$
A_1 S_1 A_2 S_2
c_1-c_2
$-c_2$
Output rate
$R_2(t)$
c_2

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 for i = 2: combine the two buffers



$$\forall t : Z_1(t) + Z_2(t) = \tilde{Z}_1(t)$$

=> outputs from the two systems are identical!

First moments of the content of buffer *i*

- Let

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

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

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

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

Correlation between the buffer contents

- Let

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

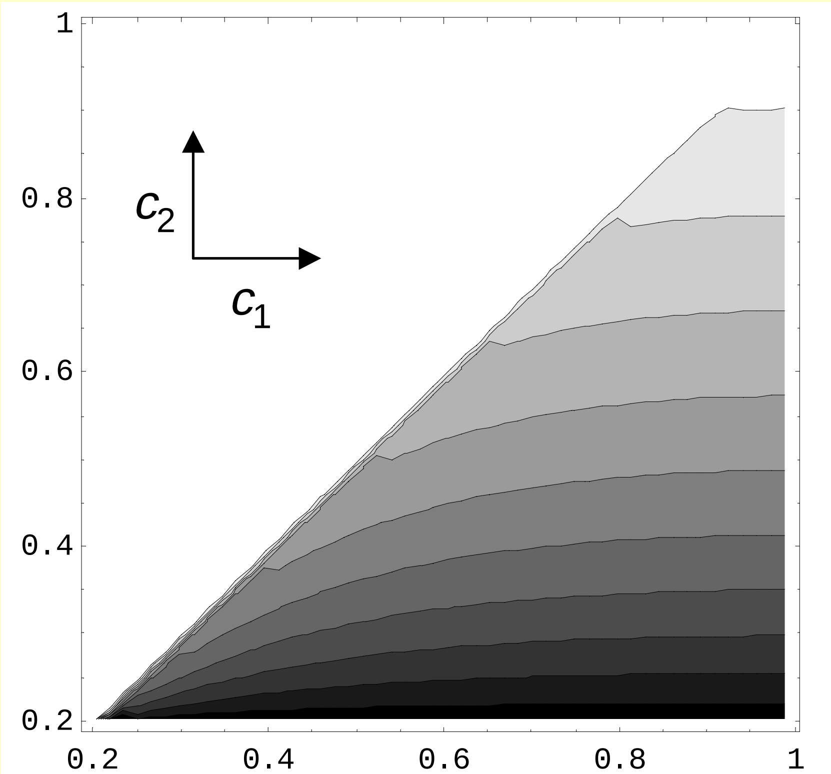
- Result: If $\rho_2 < 1$, then

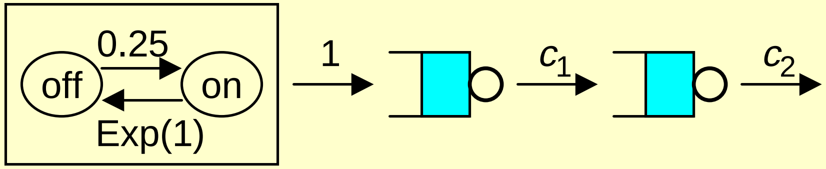
$$\mathrm{Corr}[Z_1, Z_2] = \frac{b(\kappa_0+\kappa_1)+\kappa_0^2+\kappa_0\kappa_1+\kappa_1^2}{\sqrt{(2b+2\kappa_0+\kappa_1)\kappa_1}\sqrt{(2b+2\kappa_1+\kappa_2)\kappa_2}}$$

- Idea of the proof:

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

Example (1)

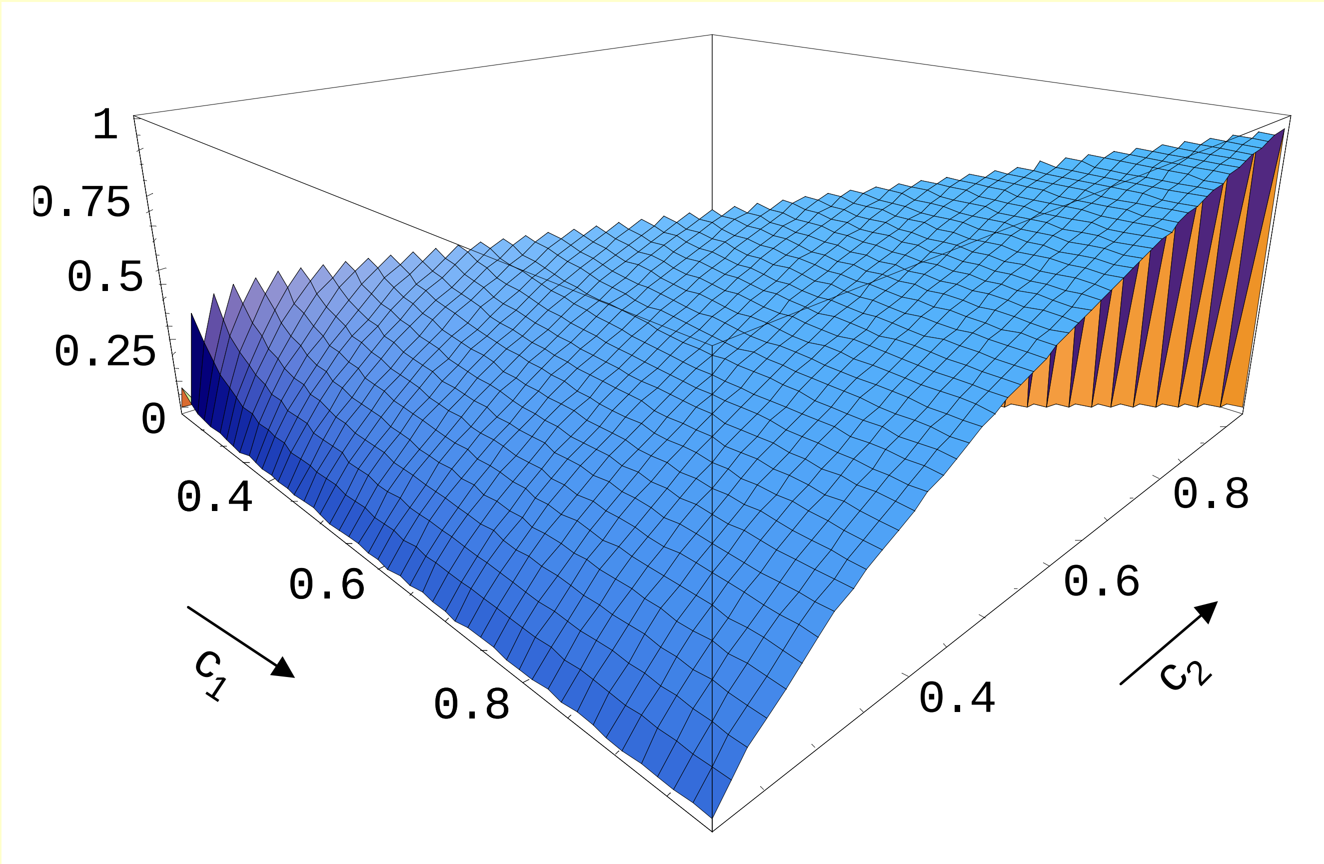




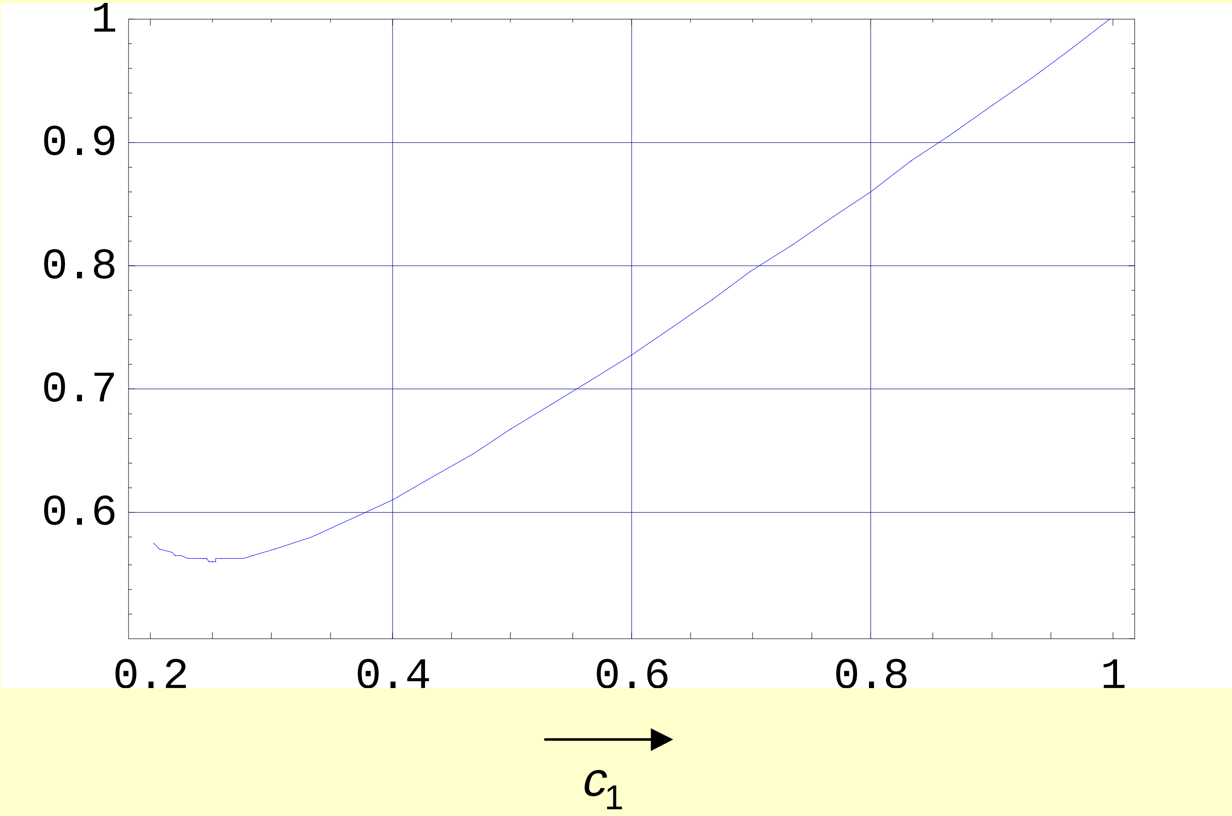
Stability condition:

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

Example (2)

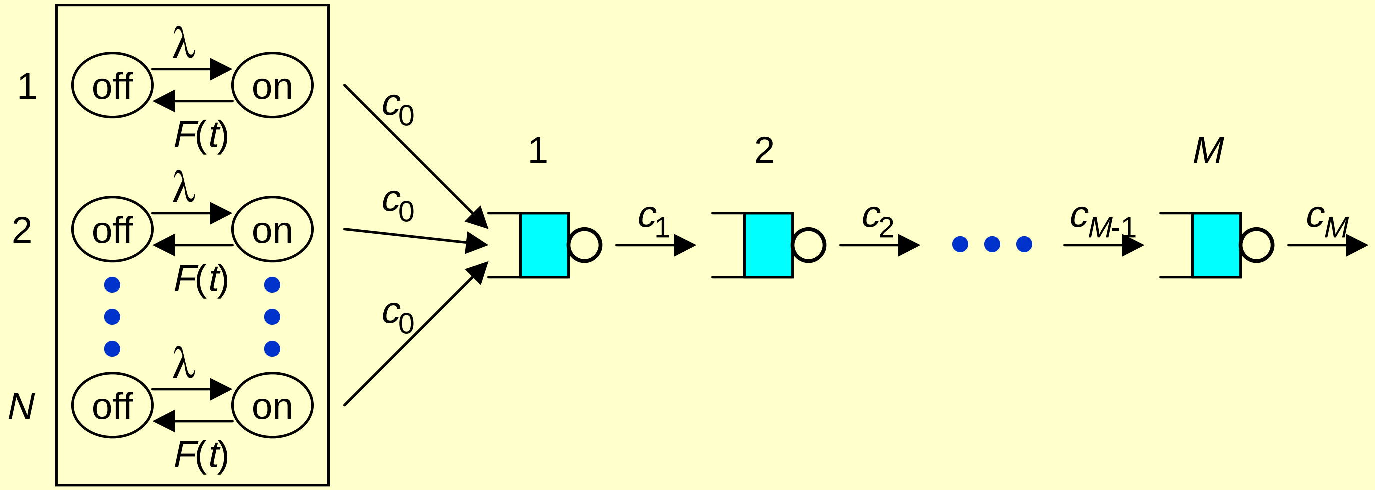


Example (3)



Generalization

- Tandem fluid queue fed by multiple homogen. on-off sources



- Assumption:

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

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

LOPPU