October 1995

UILU-ENG-95-2235
DC-169

Decision and Control

Optimal Rate Control for High Speed Telecommunication Networks

Eitan Altman and Tamer Basar

Coordinated Science Laboratory
College of Engineering
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

NSF

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UILU-ENG- DC-169	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Laboratory University of Illinois 6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION National Science Foundation
6c. ADDRESS (City, State, and ZIP Code) 1308 West Main Street Urbana, IL 61801	7b. ADDRESS (City, State, and ZIP Code) Washington, DC 20050
8a. NAME OF FUNDING / SPONSORING ORGANIZATION National Science Foundation 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Washington, DC 20050	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
OPTIMAL RATE CONTROL FOR HIGH SPEED TELECOMMUNICATION NETWORKS

12. PERSONAL AUTHOR(S)
ALTMAN, EITAN AND BAŞAR, TAMER

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	October 1995	28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Optimal rate-control, High speed networks, H^∞ control, Asynchronous Transfer Mode.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See attached

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ABSTRACT

We present in this paper two optimization-based approaches for designing controllers that dynamically regulate the rate of flow of information into a network based on feedback information about the state of the network. Such mechanisms may be used in a variety of High speed networks including in particular the Available Bit Rate service in ATM (Asynchronous Transfer Mode) networks. They result in controllers that are similar to ones that have been advocated for both end-to-end and hop-by-hop congestion control in high-speed networks. Many existing control protocols have been developed on growing available experience, using ad-hoc techniques that did not come as a result of a control-theoretical study. This is due to the high complexity of the controlled systems, that are typically decentralized, have non-linear dynamics, and may only use partial noisy delayed information. Some attempts have been made in recent years to use control theory to design flow controllers with, however, no explicit objective functions to be minimized; moreover, the class of control policies in existing theoretical work is quite restricted. The contribution of this paper is that we formulate explicitly some cost criteria to be minimized, related to performance measures such as delays, throughputs and loss probabilities. Using a linearized model, we then view the design problem as an optimal control problem. We follow two approaches to model interfering traffic and other unknown data: the H^∞ approach, and the LQG one, and determine for both cases the optimal controllers. Some simulation results complete the study.

Optimal Rate Control for High Speed Telecommunication Networks

Eitan Altman
INRIA B.P. 93
Projet Mistral
2004 Route des Lucioles
06902 Sophia Antipolis Cedex
FRANCE
altman@sophia.inria.fr

Tamer Başar*
Coordinated Science Lab.
University of Illinois
1308 West Main
Urbana, IL 61801
U.S.A.
tbasar@black.csl.uiuc.edu

October 1995

Abstract

We present in this paper two optimization-based approaches for designing controllers that dynamically regulate the rate of flow of information into a network based on feedback information about the state of the network. Such mechanisms may be used in a variety of High speed networks including in particular the Available Bit Rate service in ATM (Asynchronous Transfer Mode) networks. They result in controllers that are similar to ones that have been advocated for both end-to-end and hop-by-hop congestion control in high-speed networks. Many existing control protocols have been developed on growing available experience, using ad-hoc techniques that did not come as a result of a control-theoretical study. This is due to the high complexity of the controlled systems, that are typically decentralized, have non-linear dynamics, and may only use partial noisy delayed information. Some attempts have been made in recent years to use control theory to design flow controllers with, however, no explicit objective functions to be minimized; moreover, the class of control policies in existing theoretical work is quite restricted. The contribution of this paper is that we formulate explicitly some cost criteria to be minimized, related to performance measures such as delays, throughputs and loss probabilities. Using a linearized model, we then view the design problem as an optimal control problem. We follow two approaches to model interfering traffic and other unknown data: the H^∞ approach, and the LQG one, and determine for both cases the optimal controllers. Some simulation results complete the study.

Keywords: Optimal rate-control, High speed networks, H^∞ control, Asynchronous Transfer Mode.

---

*Research supported by Grants NSF ECS 92-20632 and NSF ECS 93-12807 from the National Science Foundation.

1 Introduction

Adaptive flow control mechanisms are being used and developed extensively in High Speed Networks, in order to guarantee high quality of services on the one hand, and allow for the efficient use of the network, on the other. The controller has to ensure that when other sources transmit, its own transmission rate is adjusted adaptively so as to avoid congestion in the network, since congestion might result in low throughputs, high delays, and high rate of losses of packets. When the interfering traffic of other sources is low, the controller is expected to allow a high rate of transmission of information so as to make the best use of the bandwidth available.

In many existing flow control protocols, some information on the state of the network is available to sources. For example, in many protocols, when a packet arrives at the destination, an acknowledgement is sent back to the source (the TCP/IP protocol in the Internet [14]). Other types of information could also be made available to the flow controller, such as information on queue length in a bottleneck node, [11], and the effective service rate available to that source in a bottleneck node [15].

Several approaches to flow control in high speed networks, based on some feedback on the state of the network, have been proposed: window flow control (e.g. [14]), credit-based control (see [16, 17] and references therein) and rate-based flow control (see [9, 17] and references therein). The last one has been selected by the ATM forum in 1994 as the main approach for flow control in the Available Bit Rate service in ATM (Asynchronous Transfer Mode); see [1].

We present in this paper two types of models for designing optimal rate based control, so as to achieve optimal performance measures: LQG (Linear-Quadratic-Gaussian) and H^{∞}. In both cases we allow for noisy delayed information. The objectives in the design of the controllers are to (1) minimize the variation of the bottleneck queue size around some desired level (2) obtain good tracking between input and output rates, and (3) minimize the jitter. The relative importance (weight) of each one of these objectives is given as data in the design of the controller. The optimal controllers that are obtained are easy to implement: they turn to be linear in the (estimated) state.

We consider a communication network with a linearized dynamics, as introduced in [2] (which extends the models in [8, 15]). We explain briefly in the next section the linearized model; the detailed derivation of the non-linear model, as well as its linear approximation is presented in [2]. Simulations further justify this linearization, see [2].

We assume that the performance measures (such as throughputs, delays and loss probabilities) are determined essentially by a bottleneck node. This assumption has both theoretical and experimental [6] justification, and is often made in the literature in order to analyze or design controllers [2, 8, 12].

The control problem is solved in Sections 3 and 4 for the case when there is no delay in the measurements, and in Section 5 when the imperfect measurements are acquired with some delay, or alternatively when the control is updated more than once during a round trip. In Section 6 we present simulation results, and some discussion on their interpretation are presented in Section 7. The paper ends with the concluding remarks of Section 8.

2 The model

We consider a discrete time model, where a time unit corresponds to a round trip delay. Let q_n denote the queue length at a bottleneck link, and μ_n denote the effective service rate available for traffic of the given source in that link at the beginning of the nth time slot. Let x_n denote the source rate during the nth time slot. We thus consider a rate-based flow control, where based on current and previous noisy information on μ_n, q_n, and x_n (to be made precise below), the controller updates the transmission rate. Hence, the queue length evolves according to

$$q_{n+1} = q_n + x_n - \mu_n. \tag{1}$$

Since several sources with varying transmission rates may share the same bottleneck link, the service rate μ_n available to the controlled source may change over time in an unpredictable way. The other sources may be represented as some interference, which we model as a stable ARMA process

$$\mu_n = \mu + \xi_n \tag{2}$$

$$\xi_n = \sum_{i=1}^{d} \ell_i \xi_{n-i} + k\phi_n \tag{3}$$

We assume that ℓ_i, which will typically be obtained by some estimation/identification procedure (using, for example, the results and framework of [10]), are such that (3) describes a stable system. The variable ϕ_n in (3) stands for disturbance, the nature of which will be described later.

We assume that, at the end of step n, the source obtains a noisy estimate of the value of μ_n, and a noisy estimate of the value of q_n. We use $\hat{\mu}_n$ to denote a measurement for the bottleneck rate, and $\hat{q}_n$ to denote a measurement for the bottleneck queue size. For our analysis, we assume that these measurements can be written as

$$\hat{\mu}_n = \mu_n + cv_n \tag{4}$$

$$\hat{q}_n = q_n + aw_n \tag{5}$$

where v_n and w_n are disturbances whose nature will be described shortly, and a and c are some constants.

In [2], the starting point was a given controller of the form

$$x_{n+1} = (1-\alpha)x_n + \alpha\hat{\mu}_n - \beta(\hat{q}_n - Q), \tag{6}$$

and the design was restricted in the choice of the parameters α and β. In particular, performance measures such as stability, expected transient (and steady state) queue lengths, the variance in steady state and the rate of convergence to steady state were studied as a function of these parameters. In (6), Q stands for some desired level of the queue length; it will appear again later in the cost function we adopt.

We define the shifted versions of the variables $x_n, q_n, \hat{\mu}_n, \hat{q}_n$:

$$\tilde{x}_n := x_n - \mu, \qquad \tilde{q}_n := q_n - Q, \qquad \hat{\tilde{\mu}}_n := \hat{\mu}_n - \mu, \qquad \hat{\tilde{q}}_n := \hat{q}_n - Q. \tag{7}$$

We further define the (scaled) variation in the input rate by bu_n, where u_n will be called the control, and b is some constant. In terms of the new variables, we now have the following dynamics:

$$\tilde{q}_{n+1} = \tilde{q}_n + \tilde{x}_n - \xi_n \tag{8}$$

$$\xi_n = \sum_{i=1}^{d} \ell_i \xi_{n-i} + k\phi_n \tag{9}$$

$$\tilde{x}_{n+1} = \tilde{x}_n + bu_n \tag{10}$$

and the observations

$$\hat{\tilde{\mu}}_n = \xi_n + cv_n, \qquad \hat{\tilde{q}}_n = \tilde{q}_n + aw_n. \tag{11}$$

We define the following quadratic cost:

$$L = g\,\|\tilde{x} - \xi\|^2 + \|\tilde{q}\|^2 + \|u\|^2 \qquad \text{where } g > 0, \quad \|u\|^2 := \sum_{n=0}^{\infty} |u_n|^2, \text{ etc...} \tag{12}$$

This definition of the cost allows us to quantify optimization criteria that have been used in previous control models [2]:

- The first term represents the quality with which the input rate tracks the available service rate, where g is a positive weighting term. A high rate of tracking is known to be desirable; see, e.g., the discussion in [8].

- The second term is a penalty for deviating from a desirable queue length. Setting some desired level of queue reflects the fact that we do not wish to have a large queue, so as to avoid losses, and on the other hand, we do not wish to have a low level of queue, since if the queue is empty and the input rate is lower than the service rate then there is a waste of potential throughput. We have not included any additional weighting on this term, as any such positive weight can be absorbed into the other variables.

- The last term stands for a penalty for high jitter, i.e. for high variability of (input) transmission rate, which is known to be undesirable; see the discussion in [8]. In many proposed telecommunication networks, the source will pay more for higher burstiness (variability) of its input rate, since highly variable input rate has typically a bad influence on other traffic. Again, as in the case of the second term, the weighting here has been normalized to 1.

We initially allow the controller to be a (measurable) function of all past and present measurements and past actions:

$$u_n = f_n(\tilde{x}_m, \hat{\tilde{\mu}}_m, \hat{\tilde{q}}_m, u_{m-1}, m = 0, \ldots, n). \tag{13}$$

Later, in Section 5, we will also consider the case where the dependence on $\hat{\tilde{\mu}}$ and $\hat{\tilde{q}}$ is with a delay of θ time units.

We consider two types of models for the disturbances:

- <u>M1: The Gaussian model</u> Here v_n, w_n, ϕ_n are taken as independent identically distributed Gaussian random variables with zero mean and unit variance. It then follows that the measurements $\hat{\mu}_n$ and $\hat{q}_n$ are unbiased. The cost to be minimized in this case is the expected value of L, written as $E[L]$.

- <u>M2: The H$^\infty$ approach</u> In this case v_n, w_n and ϕ_n are taken as unknown (deterministic) disturbances. The cost to be minimized is the ℓ_2 gain from the disturbances to the cost L, i.e.,

$$\sup_{\{\phi_n, v_n, w_n\}_{n=0}^{\infty}} \frac{L}{\|\phi, v, w\|^2} \tag{14}$$

where

$$\|\phi, v, w\|^2 := \|\phi\|^2 + \|v\|^2 + \|w\|^2 = \sum_{n=0}^{\infty} |\phi_n|^2 + |v_n|^2 + |w_n|^2 . \tag{15}$$

Let us denote the square-root of the infimum of this cost function over all admissible controllers $\{f_n\}$ by γ^*, i.e.

$$(\gamma^*)^2 = \inf_{\{f_n\}_{n=0}^{\infty}} \sup_{\{\phi_n, v_n, w_n\}_{n=0}^{\infty}} \frac{L(f; \phi, v, w)}{\|\phi, v, w\|^2}. \tag{16}$$

If there exists an actual minimizing control, say f^*, to the optimization problem above, then it can be shown [4] that, defining a soft-constrained cost function

$$L_\gamma(u; \phi, v, w) := L - \gamma^2 \|\phi, v, w\|^2$$

to be viewed as the kernel of a two player game, to be minimized by Player 1 (controlling u) and maximized by Player 2 (controlling (ϕ, v, w)), f^* has the property:

$$\sup_{\phi, v, w} L_{\gamma^*}(f^*; \phi, v, w) = \inf_f \sup_{\phi, v, w} L_{\gamma^*}(f; \phi, v, w).$$

The quantity above is the upper value of the zero-sum dynamic game with kernel L_{γ^*}, which is in fact equal to zero. It can actually be shown that for any $\gamma \geq \gamma^*$, the upper value of the game with parametrized kernel L_γ is zero, and for $\gamma < \gamma^*$, its upper value is infinite. Hence, γ^* is the smallest positive scalar γ for which the zero-sum game with kernel L_γ has a finite upper value.

Instead of obtaining f^* defined above, we will in fact solve a parametrized class of controllers, $\{f^\gamma, \gamma > \gamma^*\}$, where f^γ is obtained from

$$\sup_{\phi,v,w} L_\gamma(f^\gamma;\phi,v,w) = \inf_f \sup_{\phi,v,w} L_\gamma(f;\phi,v,w).$$

The controller f^γ will clearly have the property that it ensures a performance level γ^2 for the index adopted for M2, i.e. the attenuation is bounded by

$$\mathcal{A}(f^\gamma;\phi,v,w) := \frac{\{L(f^\gamma;\phi,v,w)\}^{1/2}}{\|\phi,v,w\|} \leq \gamma \text{ for all } \phi,v,w. \tag{17}$$

It will turn out that the limit

$$\lim_{\gamma\to\infty} f^\gamma =: f^\infty$$

is a well-defined controller, and solves uniquely the control problem with the Gaussian model, M1.

3 Complete solution to the problem with Model M2 with additional perturbation

The optimal control problem formulated above can be solved by a suitable modification of the theory of discrete-time H$^\infty$-control developed in [4]. Toward this end, we first write (8)-(10) in standard state variable form. Introduce the d-dimensional vector

$$\eta_n := (\xi_{n-d+1},\ldots,\xi_n)' \tag{18}$$

in terms of which (9) can be re-written as a first-order difference equation:

$$\eta_{n+1} = C\eta_n + k\begin{pmatrix} 0\\ .\\ .\\ .\\ 1\end{pmatrix}\phi_n \qquad \text{where} \qquad C := \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0\\ 0 & 0 & 1 & 0 & \ldots & 0\\ . & . & . & . & \ldots & \\ \ell_d & \ell_{d-1} & \ell_{d-2} & & \ldots & \ell_1\end{pmatrix} \tag{19}$$

Further introducing the $(d+2)$-dimensional vector

$$z_n := (\tilde{x}_n, \tilde{q}_n, \eta_n')' \tag{20}$$

we can write (8)-(10) in the compact form:

$$z_{n+1} = Az_n + D\phi_n + Bu_n \tag{21}$$

where

$$A := \left(\begin{array}{cc|cccc} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & -1 \\ \hline & & & & & \\ & 0 & & & C & \\ & & & & & \end{array}\right) \qquad D := \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ k \end{pmatrix} \qquad B := \begin{pmatrix} b \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \tag{22}$$

Now, the stability of the ARMA process (3) is equivalent to the stability of the matrix C in (19), which is captured in the following assumption required to hold throughout the remaining analysis:

- **A1.** All eigenvalues of C are in the unit circle in the complex plane.

Under A1, we have the following fact, which will be useful in the statement of our main result.

- **F1.** Under A1, the pair (A,B) is stabilizable.

As far as the measurements available to the controller go, we have two disturbance corrupted state measurements, given by (11), and one perfect measurement of the state, $\tilde{x}_n$. This is a "mixed" structure, that leads to a *singular* measurement equation, which is not allowed in standard theory. To circumvent this difficulty, we will first perturb $\tilde{x}_n$ with a small independent disturbance, to arrive at the measurement equation

$$\tilde{\tilde{x}}_n = \tilde{x}_n + \epsilon \tilde{\tilde{v}}_n, \qquad 0 < \epsilon << 1 \tag{23}$$

and assume that now $\tilde{\tilde{x}}_n$ is available for control purposes, instead of $\tilde{x}_n$. In the above, $\tilde{\tilde{v}}_n$ is the auxiliary disturbance introduced, and ϵ is a small positive quantity. This modification is only for mathematical convenience, which will in the end be made to vanish asymptotically by letting $\epsilon \to 0$, so as to recover the original information structure.

To introduce the modified measurement equation, let

$$y_n := (\tilde{\tilde{x}}_n, \tilde{\tilde{q}}_n, \tilde{\tilde{\mu}}_n)'. \tag{24}$$

Then

$$y_n = H z_n + E_\epsilon \tilde{v}_n \tag{25}$$

where

$$H := \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \qquad E_\epsilon := \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & c \end{pmatrix} \qquad \tilde{v}_n := \begin{pmatrix} \tilde{\tilde{v}}_n \\ w_n \\ v_n \end{pmatrix} \tag{26}$$

Since $E_\epsilon E_\epsilon' > 0$ for all $\epsilon > 0$, this is a nonsingular measurement equation. Compatible with this modification, we also modify the state equation (21), to write it as

$$z_{n+1} = A z_n + B u_n + D_\epsilon \zeta_n \tag{27}$$

where

$$D_\epsilon := \begin{pmatrix} 0 & 0 & \dots & 0 & k \\ \epsilon & 0 & \dots & 0 & 0 \end{pmatrix}' \qquad \zeta_n := (\tilde{\phi}_n, \phi_n)' \tag{28}$$

with $\tilde{\phi}_n$ being the new disturbance introduced.

To complete the description, we have to write the cost function (12) in terms of these new variables

$$L = \sum_{n=0}^{\infty} \left\{ |z_n|_N^2 + |u_n|^2 \right\} \tag{29}$$

where

$$N := \left(\begin{array}{cc|cccccc} g & 0 & 0 & . & . & . & 0 & -g \\ 0 & 1 & 0 & . & . & . & 0 & 0 \\ \hline 0 & 0 & & & & & & 0 \\ . & . & & & & & & . \\ . & . & & & 0 & & & . \\ . & . & & & & & & 0 \\ -g & 0 & 0 & . & . & . & 0 & g \end{array} \right) \tag{30}$$

and the H^∞ criterion (to be minimized) now becomes

$$\sup_{\tilde{v},\zeta} \frac{L}{\|\tilde{v},\zeta\|^2}, \tag{31}$$

the square-root of whose infimum over all admissible controllers we now denote by $\gamma^*(\epsilon)$. It turns out that $\gamma^*(\epsilon)$ is actually continuous at $\epsilon = 0$, and the quantity $\gamma^*(0)$ is equal to the γ^* defined earlier (for the unmodified problem).

Three additional properties, related to the system matrices, are now worth noting, which we state below as facts:

- **F2.** The pair (A, H) is detectable.
- **F3.** The pair (A, N) is detectable.
- **F4.** For each $\epsilon > 0$, the pair (A, D_ϵ) is controllable.

We are now in a position to state the solution to the modified H^∞ control problem formulated above, directly from Başar and Bernhard [4], Chap. 6.

Proposition 1 *Let $\epsilon > 0$ be fixed, and assumption A1 hold. Then, $\gamma^*(\epsilon)$ defined above is finite, and for all $\gamma > \gamma^*(\epsilon)$:*
1) There exists a minimal nonnegative-definite solution, M_ϵ, to

$$M = N + A' M \Lambda_\epsilon^{-1} A \tag{32}$$

such that

$$\gamma^2 I - D_\epsilon' M_\epsilon D_\epsilon > 0, \tag{33}$$

where

$$\Lambda_\epsilon := I + (BB' - \gamma^{-2} D_\epsilon D_\epsilon')M. \tag{34}$$

2) There exists a minimal nonnegative-definite solution, Σ_ϵ, to

$$\Sigma = D_\epsilon D_\epsilon' + A\Sigma R_\epsilon^{-1} A' \tag{35}$$

such that

$$\gamma^2 I - N^{1/2} \Sigma_\epsilon N^{1/2} > 0, \tag{36}$$

where

$$R_\epsilon := I + \left(H'(E_\epsilon E_\epsilon')^{-1} H - \gamma^{-2} N \right) \Sigma. \tag{37}$$

3)

$$\Sigma_\epsilon^{1/2} M_\epsilon \Sigma_\epsilon^{1/2} < \gamma^2 I. \tag{38}$$

4) An H^∞-controller that ensures the bound γ^2 in the index (31) is

$$u_{n,\epsilon}^\gamma = f_\epsilon^\gamma(\hat{z}_{n|n}) = -B' M_\epsilon \Lambda_\epsilon^{-1} A \hat{z}_{n|n}, \qquad n = 0, 1, \ldots \tag{39}$$

where $\hat{z}_{n|n}$ is generated by

$$\hat{z}_{n|n} = \left(I + \Sigma_\epsilon H'(E_\epsilon E_\epsilon')^{-1} H - \gamma^{-2} \Sigma_\epsilon M_\epsilon \right)^{-1} \left(\check{z}_n + \Sigma_\epsilon H'(E_\epsilon E_\epsilon')^{-1} y_n \right) \tag{40}$$

$$\check{z}_{n+1} = A\check{z}_n + Bu_n + A\Sigma_\epsilon R_\epsilon^{-1} \left[\gamma^{-2} N \check{z}_n + H'(E_\epsilon E_\epsilon')^{-1} (y_n - H\check{z}_n) \right] \tag{41}$$

5) The controller (39) leads to a bounded input-bounded state stable system, i.e. for all bounded disturbances $\tilde{v}, \zeta$, the system state z_n and the filter state $\hat{z}_{n|n}$ remain bounded. Equivalently, the two matrices

$$(I - BB' M_\epsilon \Lambda_\epsilon^{-1}) A \quad \textit{and} \quad \left(I - H'(E_\epsilon E_\epsilon')^{-1} H \Sigma_\epsilon R_\epsilon^{-1} \right) A'$$

have all their eigenvalues in the unit circle.

If any one of the three properties (1)-(3) above does not hold, then $\gamma < \gamma^*(\epsilon)$. ◇

Remark 1 The results of Proposition 1 are valid not only for all finite γ's, larger than $\gamma^*(\epsilon)$, but also as $\gamma \to \infty$. In this limiting case, which captures the solution to ϵ-perturbed problem under the Gaussian model (M1) of the disturbance [5], the constraints (33), (36) and (37) become irrelevant, and all other expressions admit "well-defined" forms, obtained by simply setting $\gamma^{-2} = 0$. ◇

4 The limiting solution as $\epsilon \downarrow 0$

We next study the limit of the solution given in Proposition 1 as $\epsilon \downarrow 0$. For the limit to be well defined, it will be sufficient to show that equations and relationships (32)-(38) are well-defined in the limit as $\epsilon \downarrow 0$.

The first three depend continuously on $\epsilon \geq 0$, and hence in the limit (32) will be replaced by

$$M = N + A'M\Lambda_0^{-1}A \tag{42}$$

where

$$\Lambda_0 := I + \text{diag}(b^2, 0, \ldots, 0, -\gamma^2 k^2)M \tag{43}$$

and (42) will have to be solved under the scalar condition (as counterpart of (33)):

$$k^2(0, \ldots, 0, 1)M(0, \ldots, 0, 1)' < \gamma^2 \tag{44}$$

which guarantees invertibility of (43). Again, for each $\gamma > \gamma^*(0)$, (42) admits a unique minimal nonnegative-definite solution, which also satisfies (44).

The limiting analysis of (35)-(37) is somewhat more complicated, since it involves the inverse of a matrix that becomes singular at $\epsilon = 0$ (which is $E_\epsilon E_\epsilon'$). However, this isolated singularity does not lead to any singularity for the solution of (35) as to be shown below. We will in fact show that in the limit as $\epsilon \downarrow 0$, the solution of (35) will be in the form

$$\Sigma_0 = \begin{pmatrix} 0 & . & . & . & 0 \\ . & & & & \\ . & & P & & \\ . & & & & \\ 0 & & & & \end{pmatrix} \tag{45}$$

where $P \geq 0$ is of dimension $(d+1) \times (d+1)$. To see this, let us first rewrite ΣR_ϵ^{-1} as

$$\Sigma R_\epsilon^{-1} = \Sigma^{1/2} \left[I + \Sigma^{1/2} H'(E_\epsilon E_\epsilon')^{-1} H \Sigma^{1/2} - \gamma^{-2} \Sigma^{1/2} N \Sigma^{1/2} \right]^{-1} \Sigma^{1/2}$$

where $\Sigma^{1/2}$ denotes the unique nonnegative definite square root of Σ. Let the $(d+1) \times (d+1)$-dimensional lower block of N be denoted by $\tilde{N}$, which is

$$\tilde{N} = \text{diag}(1, 0, \ldots, 0, g) \tag{46}$$

and suppose that Σ is given by (45), where P is arbitrary. Then, straightforward manipulations lead to

$$\Sigma^{1/2} N \Sigma^{1/2} = \text{block-diag}\left(0,\ P^{1/2} \tilde{N} P^{1/2}\right)$$

$$\Sigma^{1/2} H'(E_\epsilon E_\epsilon')^{-1} H \Sigma^{1/2} = \text{block-diag}\left(0,\ P^{1/2} \tilde{H}'(\tilde{E}\tilde{E}')^{-1} \tilde{H} P^{1/2}\right)$$

where

$$\tilde{H} := \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}}_{2\times(d+1)} \qquad \tilde{E} := \underbrace{\begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}}_{2\times 2} \tag{47}$$

Hence,

$$\begin{aligned}
\Sigma R_\epsilon^{-1} &= \Sigma^{1/2}\left[\text{block-diag}\left(1,\ I + P^{1/2}\left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P^{1/2}\right)\right]^{-1}\Sigma^{1/2} \\
&= \text{block-diag}\left(0,\ P^{1/2}\left[I + P^{1/2}\left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P^{1/2}\right]^{-1}P^{1/2}\right) \\
&= \text{block-diag}(0,\ \tilde{S})
\end{aligned} \tag{48}$$

which is independent of ϵ. Pre- and post-multiplying this expression by A and A', we obtain

$$A\Sigma R_\epsilon^{-1}A' = \text{block-diag}\left(0,\ \tilde{A}P^{1/2}\left[I + P^{1/2}\left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P^{1/2}\right]^{-1}P^{1/2}\tilde{A}'\right)$$

where $\tilde{A}$ is the $(d+1)\times(d+1)$ dimensional lower block of A, i.e.

$$\tilde{A} := \left(\begin{array}{c|c} 1 & 0\ .\ .\ .\ 0\ \text{-}1 \\ \hline 0 & \\ . & \\ . & C \\ . & \\ 0 & \end{array}\right) \tag{49}$$

Furthermore, since

$$\lim_{\epsilon\to 0} D_\epsilon D_\epsilon' = \text{diag}(0,\dots,0,k^2),$$

it readily follows that the structure (45) is consistent with (35). Introducing a $(d+1)$-dimensional vector $\tilde{D}$,

$$\tilde{D} := (0,\dots,0,k)', \tag{50}$$

the equation to be satisfied by P can now be written as

$$P = \tilde{D}\tilde{D}' + \tilde{A}P\left[I + \left(\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}\tilde{N}\right)P\right]^{-1}\tilde{A}', \tag{51}$$

and (36) becomes equivalent to

$$\gamma^2 I - \tilde{N}^{1/2}P\tilde{N}^{1/2} > 0. \tag{52}$$

Finally, (38) simplifies to

$$P^{1/2}\tilde{M}P^{1/2} < \gamma^2 I \tag{53}$$

where $\tilde{M}$ is the $(d+1) \times (d+1)$-dimensional lower block of M.

Hence, (32)-(34) are well-defined at $\epsilon = 0$, and (35)-(38) admit well-defined limits in the space of dimension $(d+1)$ instead of $(d+2)$.

To complete the limiting analysis, we still have to find the limiting form of the controller (39)-(41). Toward this end, write z, as $\epsilon \downarrow 0$, as

$$z = (\alpha, \beta')' \tag{54}$$

where α is of dimension 1, and β of dimension $d+1$. Compatibly, write (again as $\epsilon \downarrow 0$):

$$\hat{z}_{n|n} = (\hat{\alpha}_{n|n},\ \hat{\beta}'_{n|n})'; \qquad \breve{z}_n = (\breve{\alpha}_n,\ \breve{\beta}'_n)'. \tag{55}$$

Then, using the earlier manipulations, and some straightforward extensions, it can be shown that $\hat{\alpha}_{n|n}, \hat{\beta}_{n|n}, \breve{\alpha}_n, \breve{\beta}_n$ are generated by (from (40)-(41), and using the notation introduced by (46)-(50)):

$$\hat{\alpha}_{n|n} = \breve{\alpha}_n \tag{56}$$

$$\breve{\alpha}_{n+1} = \breve{\alpha}_n + bu_n \tag{57}$$

$$\hat{\beta}_{n|n} = \left(I + P\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H} - \gamma^{-2}P\tilde{M}\right)^{-1}\left(\breve{\beta}_n + P\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{y}_n\right) \tag{58}$$

$$\breve{\beta}_{n+1} = \tilde{A}\breve{\beta}_n + \tilde{A}\tilde{S}\left[\gamma^{-2}\tilde{N}\breve{\beta}_n + \tilde{H}'(\tilde{E}\tilde{E}')^{-1}(\tilde{y}_n - \tilde{H}\breve{\beta}_n)\right] \tag{59}$$

$$\tilde{y}_n := (\hat{\tilde{q}}_n, \hat{\tilde{\mu}}_n)' \tag{60}$$

Clearly, (57) generates $\tilde{x}_n$ whose perfect value is already available to the controller. Hence, the limiting controller is of dimension $d+1$, with the worst-case observer given by (58)-(59). The counterpart of (39) is then

$$u_n^{\gamma} = f^{\gamma}(x_n, \hat{\beta}_{n|n}) = -B'M_0\Lambda_0^{-1}A\begin{pmatrix}\tilde{x}_n \\ \hat{\beta}_{n|n}\end{pmatrix}, \qquad n = 0, 1, \ldots \tag{61}$$

This result is summarized in the following Theorem:

Theorem 1 *Consider the original control problem formulated in Section 2, under the H^∞ model* (M2) *of the disturbance. Let Assumption A1 hold. Then γ^* defined by (16) is finite, and for all $\gamma > \gamma^*$:*
1) There exists a minimal nonnegative-definite solution, M_0, to (42) such that (44) holds.
2) There exists a minimal nonnegative-definite solution, P, to (51), such that (52) holds.
3) Condition (53) holds.
4) An H^∞-controller that ensures the bound γ^2 in the index (14) is given by (61).
5) The controller (61) leads to a bounded input-bounded state stable system (see Proposition 1 (5)).
For a given $\gamma > 0$, if any one of the properties (1)-(3) above does not hold, then $\gamma < \gamma^*$. $\diamond$

We immediately have the following corollary to Theorem 1, obtained by letting $\gamma \to \infty$, which solves the original problem under the Gaussian interpretation (M1) of the disturbances (see Remark 1).

Corollary 1 *Under A1, the optimal control problem with the disturbance model M1 admits the unique solution:*

$$u_n^\infty = f^\infty(x_n, \hat{\beta}_{n|n}) = -B'\overline{M}(I + BB'\overline{M})^{-1} A \begin{pmatrix} \tilde{x}_n \\ \hat{\beta}_{n|n} \end{pmatrix}, \qquad n = 0, 1, \ldots \tag{62}$$

where $\overline{M}$ is the unique nonnegative-definite solution of

$$\overline{M} = N + A'\overline{M}(I + BB'\overline{M})^{-1} A \tag{63}$$

in the class of nonnegative-definite matrices, and

$$\hat{\beta}_{n|n} = \left(I + \overline{P}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H}\right)^{-1} \left(\breve{\beta}_n + \overline{P}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{y}_n\right) \tag{64}$$

$$\breve{\beta}_{n+1} = \tilde{A}\breve{\beta}_n + \tilde{A}\overline{S}\tilde{H}'(\tilde{E}\tilde{E}')^{-1}(\tilde{y}_n - \tilde{H}\breve{\beta}_n) \tag{65}$$

where

$$\overline{S} := \overline{P}\left[I + \tilde{H}'(\tilde{E}\tilde{E}')^{-1}\tilde{H}\overline{P}\right]^{-1} \tag{66}$$

and $\overline{P}$ is the unique nonnegative-definite solution of

$$\overline{P} = \tilde{D}\tilde{D}' + \tilde{A}\overline{S}\tilde{A}' \tag{67}$$

in the class of nonnegative definite matrices.
Furthermore, the controller (62) leads to a stable system. ◇

5 Delayed measurements

In our basic model, presented in Section 2, we considered a discrete-time controlled system, where a time unit corresponds to a round trip delay. In applications where the end-to-end delays are quite large, it might be highly inefficient to wait for a whole round trip delay till the control is updated. In such cases, one may choose one of the following two options:
(i) Consider some intermediate nodes as virtual sources and destinations, and use a hop-by-hop flow control. In that case, one may use at each hop the optimal control schemes proposed in the previous sections.
(ii) Allow for updating of the control several times (say $\theta \geq 1$ times) during a round trip. In that case, a basic time unit corresponds to the round trip time divided by $(1 + \theta)$.

One may, of course, also use a combination of the two possibilities above, and allow for a hop-by-hop control where, at each hop, control is updated at times which are shorter than the round trip hop time.

We shall focus in this section on the second possibility above, and design the corresponding optimal controllers. The system can still be modeled by the dynamics described in Section 2. However, since the basic time unit is now shorter (by a factor of $1+\theta$) than the time unit considered in the previous sections, the basic parameters of the model have to be rescaled accordingly. For example, μ, appearing in (2), will now be one-$(\theta+1)th$ of the earlier one, and so will μ_n, x_n and u_n. In terms of the new time unit, the problem formulation will then be exactly as the one in Section 2 – with one major difference. Now the measurements $\hat{\tilde{\mu}}$ and $\hat{\tilde{q}}$ are acquired (for control purposes) with a delay of $\theta \geq 1$ time units. Hence, the controller (13) will be of the form

$$u_n = f_n(\tilde{x}_m, \hat{\tilde{\mu}}_{m-\theta}, \hat{\tilde{q}}_{m-\theta}, u_{m-1}, m = 0, \ldots, n), \tag{68}$$

where the convention (throughout) is to take $\hat{\tilde{\mu}}_{m-\theta}$ and $\hat{\tilde{q}}_{m-\theta}$ to be identically *zero* for $m < \theta$. Here we could also have taken the dependence on $\tilde{x}_m$ to be with a delay of θ units, but since $\tilde{x}_m$ is generated by noise-free dynamics, this would not make any difference in the end result; that is, any performance that is achievable using $\tilde{x}_m$ could also be achieved using $\tilde{x}_{m-\theta}$.

We note that even though this structure was arrived at by increasing the frequency of control updates and by rescaling the parameters of the model (as discussed above), it can also be interpreted for a fixed time scale as the control having access to measurements with a delay of θ time units. In the framework of this interpretation, we denote the minimum H$^\infty$ performance level $\gamma^*(\epsilon)$ by $\gamma^*_\theta(\epsilon)$, to show explicitly its dependence on the delay factor θ. When we vary θ (as we will do in the sequel) this will be done precisely in this framework, with the time-scale kept fixed. This now brings us to the point where we present the counterpart of Proposition 1 for the delayed measurements case. The small-noise perturbed problem is precisely the one formulated in Section 3, with the only difference being that now the control has access to $y_{m-\theta}, m \leq n$, at time n.

Proposition 2 *Consider the framework of Section 3, but with the controller having the structural form discussed above, with $\theta \geq 1$. Let $\epsilon > 0$ be fixed, and assumption A1 hold. Then:*
1) $\gamma^*_\theta(\epsilon)$ is finite, and for each fixed $\epsilon > 0$, the ordering $\theta_2 > \theta_1 \geq 1$ implies that

$$\gamma^*_{\theta_2}(\epsilon) \geq \gamma^*_{\theta_1}(\epsilon) \geq \gamma^*(\epsilon),$$

where $\gamma^*(\epsilon)$ is as defined in Proposition 1.
2) An H$^\infty$-controller that ensures a bound $\gamma^2 > (\gamma^*_\theta(\epsilon))^2$ in the index (31) is

$$u^\gamma_{n,\epsilon,\theta} = f^\gamma_{\epsilon,\theta}(\hat{z}_{n|n-\theta}) = -B'M_\epsilon\Lambda_\epsilon^{-1}A\left(I - \gamma^{-2}\tilde{\Sigma}_{\epsilon,\theta}M_\epsilon\right)^{-1}\hat{z}_{n|n-\theta}, \qquad n = 0, 1, \ldots \tag{69}$$

where $\hat{z}_{n|n-\theta} = \xi^{(n)}_n$, with $\xi^{(n)}_n$ being the last step of the iteration $\{\xi^{(n)}_k\}^n_{k=n-\theta+1}$ generated by

$$\xi^{(n)}_{k+1} = A\left(I - \gamma^{-2}\tilde{\Sigma}_{\epsilon,\theta}N\right)^{-1}\xi^{(n)}_k + Bu_k\,; \quad \xi^{(n)}_{n-\theta+1} = \begin{cases} \check{z}_{n-\theta+1} & n \geq \theta \\ 0 & else \end{cases}, \tag{70}$$

and $\{\check{z}_n\}_{n=0}^{\infty}$ *is as generated by (41). The matrix* $\tilde{\Sigma}_{\epsilon,\theta}$ *is the* θ*-th term generated by the recursion*

$$\tilde{\Sigma}_{\epsilon,k+1} = A\tilde{\Sigma}_{\epsilon,k}\left(I - \gamma^{-2}N\tilde{\Sigma}_{\epsilon,k}\right)^{-1}A' + D_\epsilon D_\epsilon'; \quad \tilde{\Sigma}_{\epsilon,1} = \Sigma_\epsilon, \tag{71}$$

satisfying the two spectral radius conditions

$$N^{1/2}\tilde{\Sigma}_{\epsilon,k}N^{1/2} < \gamma^2 I, \quad k = 1,\ldots,\theta \tag{72}$$

$$M_\epsilon^{1/2}\tilde{\Sigma}_{\epsilon,k}M_\epsilon^{1/2} < \gamma^2 I, \quad k = 1,\ldots,\theta+1. \tag{73}$$

The initializing matrix Σ_ϵ *in (71) is a minimal nonnegative-definite solution of (35), satisfying (36), and all other terms, such as* M_ϵ*,* Λ_ϵ*,* D_ϵ*, are as defined in Proposition 1.*

If any one of the conditions of Proposition 1, or (72)-(73) do not hold, then $\gamma < \gamma_\theta^*(\epsilon)$*.*

Proof. The first statement above follows from the fact that in H^∞-optimal control, less on-line information on the uncertainties cannot lead to improved optimum performance, which in this case is quantified by $\gamma_\theta^*(\epsilon)$. The other statements of the Proposition follow directly from Theorem 6.6 (p. 276) of the 2nd edition of [4]. Even though this Theorem states the solution of only the finite-horizon case, its extension to the infinite-horizon case (which is dealt with here) follows by mimicking the arguments used in the non-delay case; see Theorem 5.5, p. 226 of the 2nd edition of [4]. ◇

Remark 2 The ordering of the γ^*'s in part 1 of Proposition 2 is valid, as discussed prior to the statement of the Proposition, under the interpretation that the round-trip delay is fixed, and θ reflects the delay in the acquisition of the measurements. The problem of relevance, however, is the one where the number of control updates is increased (from 1 to $\theta+1$) during each round trip. Under this interpretation, the time scaling will not be fixed, and we would then expect the H^∞ performance level to be a nonincreasing function of θ (instead of being nondecreasing, as in the case of Proposition 2), because the control is applied more frequently on a given time interval. ◇

To present the counterpart of Theorem 1 for the $\theta \geq 1$ case, we have to study the limiting process as $\epsilon \downarrow 0$. The analysis of Section 4 applies here intact, with the exception of the control law; we also have to determine the limiting expression for $\tilde{\Sigma}_{\epsilon,k}$.

Compatible with that of $\Sigma_0 = \lim_{\epsilon\downarrow 0}\Sigma_\epsilon$, it is not difficult to show that

$$\tilde{\Sigma}_{0,k} = \text{ block-diag }\left(0,\ \tilde{P}_k\right) \tag{74}$$

where the $(d+1)$-dimensional square matrix $\tilde{P}_k$ is generated by the recursion

$$\tilde{P}_{k+1} = \tilde{A}\tilde{P}_k\left(I - \gamma^{-2}\tilde{N}\tilde{P}_k\right)^{-1}\tilde{A}' + \tilde{D}\tilde{D}'; \quad \tilde{P}_1 = P, \tag{75}$$

where P is a minimal nonnegative-definite solution of (51), satisfying (52), and $\tilde{A}$, $\tilde{D}$ are as defined by (49) and (50), respectively. In view of this simplification, the two spectral radius conditions (72) and (73) become, respectively

$$\tilde{N}^{1/2}\tilde{P}_k\tilde{N}^{1/2} < \gamma^2 I\,, \quad k = 1,\ldots,\theta \tag{76}$$

$$\tilde{M}^{1/2}\tilde{P}_k\tilde{M}^{1/2} < \gamma^2 I\,, \quad k = 1,\ldots,\theta+1\,, \tag{77}$$

where $\tilde{N}$ was defined by (46) and $\tilde{M}$ in conjunction with (53).

To obtain the limiting form of the controller, we write (compatibly with (54) and (55)):

$$\hat{z}_{n|n-\theta} = \left(\hat{\alpha}_{n|n-\theta},\, \hat{\beta}'_{n|n-\theta}\right)'\,; \qquad \xi_k^{(n)} = \left(\tilde{\alpha}_k^{(n)},\, \tilde{\beta}_k^{(n)'}\right)' \tag{78}$$

Then, some manipulations yield:

$$\begin{array}{l} \tilde{\alpha}_{k+1}^{(n)} = \tilde{\alpha}_k^{(n)} + bu_k\,; \quad \tilde{\alpha}_{n-\theta+1}^{(n)} = \left\{ \begin{array}{cc} \tilde{x}_{n-\theta+1} & n \geq \theta \\ 0 & \text{else} \end{array} \right. \\ \Rightarrow \quad \tilde{\alpha}_k^{(n)} = \tilde{x}_k\,, \quad k = 0,1,\ldots\ \forall\, n \geq 0 \qquad \Rightarrow \qquad \hat{\alpha}_{n|n-\theta} = \tilde{x}_n \end{array} \tag{79}$$

and

$$\begin{array}{c} \hat{\beta}_{n|n-\theta} = \tilde{\beta}_n^{(n)} \\ \tilde{\beta}_{k+1}^{(n)} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}}_{(d+1)\times(d+2)} A\left(I - \gamma^{-2}\tilde{\Sigma}_{0,\theta}N\right)^{-1} \begin{pmatrix} \tilde{x}_k \\ \tilde{\beta}_k^{(n)} \end{pmatrix}\,; \\ \tilde{\beta}_{n-\theta+1}^{(n)} = \left\{ \begin{array}{cc} \breve{\beta}_{n-\theta+1} & n \geq \theta \\ 0 & \text{else} \end{array} \right. \end{array} \tag{80}$$

where $\breve{\beta}_n$, $n \geq 1$ is generated by (59). Note that in the delayed measurement case we do not have a separation between $\hat{\alpha}$ and $\hat{\beta}$ as clean as in the case with $\theta = 0$. The dynamics for $\tilde{\beta}^{(n)}$ above, and thereby that of $\hat{\beta}_{n|n-\theta}$ use also $\tilde{x}$ as an input, whereas (58) and (59) did not.

In view of these limiting expressions, the limiting expression for the controller (69), as $\epsilon \downarrow 0$, is

$$u_{n,\theta}^{\gamma} = f_\theta^{\gamma}(x_n, \hat{\beta}_{n|n-\theta}) = -B'M_0\Lambda_0^{-1}A\left(I - \gamma^{-2}\tilde{\Sigma}_{0,\theta}M_0\right)^{-1} \begin{pmatrix} \tilde{x}_n \\ \hat{\beta}_{n|n-\theta} \end{pmatrix}, \qquad n = 0,1,\ldots \tag{81}$$

This now leads to the following Theorem:

Theorem 2 *Consider the original control problem formulated in Section 2, under the* H^∞ *model* (M2) *of the disturbance, and with delayed measurements, with the optimum* H^∞ *performance level denoted by* γ_θ^*. *Let Assumption A1 hold. Then* γ_θ^* *is finite, and for all* $\gamma > \gamma_\theta^*$:
1) Statements 1)-3) of Theorem 1 hold.
2) Conditions (76) and (77) hold, and the matrix sequence generated by (75) is well defined and nonnegative definite.
3) An H^∞*-controller that ensures the bound* γ^2 *in the index (14) under* θ*-delayed measurements is given by (81).*
For a given $\gamma > 0$, *if any one of the conditions above does not hold, then* $\gamma < \gamma_\theta^*$. ◇

By letting $\gamma \to \infty$ in the theorem above, we arrive at the following counterpart of Corollary 1 under delayed measurements.

Corollary 2 *Under assumption A1, the optimal control problem with the disturbance model M1, and* θ*-delayed measurements, admits the unique solution:*

$$u_{n,\theta}^{\infty} = f_\theta^\infty(x_n, \hat{\beta}_{n|n-\theta}) = -B'\overline{M}(I + BB'\overline{M})^{-1}A\begin{pmatrix} \tilde{x}_n \\ \hat{\beta}_{n|n-\theta} \end{pmatrix}, \qquad n = 0, 1, \ldots \tag{82}$$

where $\overline{M}$ *is the unique nonnegative-definite solution of (63) in the class of nonnegative-definite matrices, and*

$$\hat{\beta}_{n|n-\theta} = \tilde{\beta}_n^{(n)}; \qquad \tilde{\beta}_{k+1}^{(n)} = \tilde{A}\tilde{\beta}_k^{(n)}, \; \tilde{\beta}_{n-\theta+1}^{(n)} = \begin{cases} \breve{\beta}_{n-\theta+1} & n \geq \theta \\ 0 & else \end{cases} \tag{83}$$

where $\breve{\beta}_n$, $n \geq 1$, *is generated by (65), and* $\tilde{A}$ *is as defined by (49).* ◇

6 Simulation Results

We present in this section some simulation results using both the H^∞ and the Gaussian model, and with no delay in the measurements (or with control updated only once during each round trip); hence, this will be an illustration of the results of Section 4. For simulation purposes, the observation noises v_n and w_n are chosen as standard i.i.d. Gaussian sequences. In order to illustrate the effectiveness of the H^∞ (worst-case) controller in the context of telecommunications, we present results obtained under different choices of perturbations (interfering traffic) ϕ_n (in (3)):
(i) highly correlated perturbations: $\phi_n = \sin(0.2n)$,
(ii) i.i.d. Gaussian perturbations,
(iii) i.i.d. uniformly distributed perturbations (over $[-1, 1]$).

Below we illustrate how the design objectives influence the behavior of the system when driven by the optimal controller. More precisely, we shall focus on the influence of the parameters g (see (12)) and b (see (10)) on the system.

The parameter g will influence the tracking between input and output rates. This is seen from equations (8) and (12). The larger g is, the more costly the deviations between the input and output rates will be. Indirectly, this may result in reducing the variations of the queue size (see (8)).

The parameter b is expected to influence the jitter, i.e. the variability of the transmission rate $\tilde{x}$, which is given by (see (10)):

$$\hat{u}_n := bu_n = \tilde{x}_{n+1} - \tilde{x}_n. \tag{84}$$

It is easily seen from (10) and (12) that choosing a smaller value of b results in a higher weight on deviations of u from zero, and thus a higher weight on the variability of $\tilde{x}$. Thus, we would expect that smaller values for b will be useful when designing controllers for video or voice traffic, where the transmission rate is typically required to be regular (i.e. low jitter). (Note that one may consider $\hat{u}_n$ as the control, instead of u_n; then the cost L in (12) becomes $L = g\|\tilde{x} - \xi\|^2 + \|\tilde{q}\|^2 + b^{-2}\|\hat{u}\|^2$. This shows that indeed lower values of b result in higher weight on the jitter.)

Finally, if g is small and b is large, then the second term in (12) becomes important, which means that we pay more for deviations of the queue length from its target value Q. In practice this part will be responsible for decreasing loss probabilities (that occur when queue sizes are high) and increasing throughputs (since we may lose potential throughput when the queue is empty, if, at that time, the input rate is lower than the output rate).

We shall examine three situations in the simulations below:

Case 1: The three different criteria are weighted equally in the cost L to be minimized (see (10) and (12)): $g = b = 1$.
Case 2: $b = g = 0.1$. We thus place relatively less emphasis on the tracking between input and output rates and thus on the variability of the queue length; the objective that is most weighted is that of minimizing the jitter $\hat{u}$.
Case 3: $b = 10, g = 0.1$. The objective that is most weighted is that of minimizing the deviations of the queue size from Q.

To describe the interfering traffic, we have chosen the same parameter values as in [2]: an ARMA model of order two with $\ell_1 = 0.7, \ell_2 = -0.3$, $k = \sqrt{5}$. The parameters a and c (appearing in (4)-(5)) were chosen to be $\sqrt{2}$. We chose $Q = 30$ and $\mu = 40$ units. (A unit corresponds to the number of packets transmitted during a round trip delay). The duration of each simulation was 140 time units. The queues were initially empty. A steady state was reached in each case after 20 time units. The design of the controller is for a steady state behavior. We present in the table below the performance (attenuation) for a transient period (first 100 units) as well as the steady state period (last 100 units). In all cases, the performance is indeed considerably better at the steady state.

For each of the cases above, we considered both of the scenarios below:

- an almost optimal H^∞ controller, with γ very close to (slightly larger than) the optimal (smallest) value γ^* (see (16)), so as to operate under stable conditions.

- very high values of γ (of the order of 100), which resulted in a controller that is optimal for the Gaussian model (M1).

For Case 1, the optimal (smallest) value of γ (see (16)) for the H^∞ controller is $\gamma^* = 5.9$. We chose $\gamma = 6$. For Case 2 we obtained $\gamma^* = 13.5$ and chose $\gamma = 13.8$, and for Case 3 we obtained $\gamma^* = 3.7$ and chose $\gamma = 3.8$.

For each combination of the parameter values, we present here 4 curves: the evolution of the queue size, the input versus output rate over the entire simulation interval as well as over a typical shorter period (a zoomed-in curve), and the evolution of the control. (Multiplying the control evolution by the parameter b provides us with the jitter. The scales for different control curves were adjusted so that when multiplied by b they result in the same scaling for $\hat{u}$.)

	Parameters					*Simulation Results*	
Sim. number	γ	b	g	*pert.*	obs. noise	s.s. attenuation	tr. attenuation
1	100	0.1	0.1	sin	Gauss.	9.17	9.91
2	13.8	0.1	0.1	sin	Gauss.	7.01	8.28
3	100	1	1	sin	Gauss.	2.78	4.42
4	6	1	1	sin	Gauss.	3.40	4.79
5	100	10	0.1	sin	Gauss.	1.71	3.54
6	3.8	10	0.1	sin	Gauss.	1.83	3.46
7	100	1	1	Gauss.	Gauss.	2.99	4.14
8	6	1	1	Gauss.	Gauss.	3.14	4.17
9	100	0.1	0.1	Gauss.	Gauss.	3.67	5.70
10	13.8	0.1	0.1	Gauss.	Gauss.	5.10	6.28
11	100	1	1	Unif.	Unif.	2.48	6.03
12	6	1	1	Unif.	Unif.	2.63	6.02
13	100	0.1	0.1	Unif.	Unif.	3.56	7.67
14	13.8	0.1	0.1	Unif.	Unif.	4.77	8.43

7 Conclusions drawn from the simulations

As expected, the curves below indicate that smaller b and g result in low jitter. Indeed, the jitter (bu_n) is the smallest in Figures 1, 2, 9, 10, 13 and 14, for which $g = b = 0.1$; its values are between -1 and +1. In Figures 3, 4, 7, 8, 11 and 12, in which $g = b = 1$, the jitter has much higher oscillations, of amplitude of 3. In Figures 5 and 6, in which $g = 0.1, b = 10$, it has oscillations of amplitude 5. (Note that for a fair comparison of all these figures, u has to be multiplied by b).

Large b and relatively small g result in the lowest queue lengths. Indeed, in Figures 5, 6, for which $g = 0.1, b = 10$, the amplitude of the oscillations of the queue length is 5, whereas in Figure 1 (g=b=0.1) the oscillations are of amplitude 20.

Finally, relatively large values of g lead indeed to much better tracking of the input with respect to the output rate, as can be seen from Figures 3, 4, 7, 8, 11, 12, where $g = b = 1$. We observe from the figures that relatively large values of g also lead to smaller variation in the queue length around its target level.

An interesting question here is whether in specific applications one should choose an optimal H^∞ controller (designed for the worst case behavior of noise and perturbations) or a Gaussian one. When both the noise and the perturbations are Gaussian, we see from the Table above (Simulations 7-10) that, as expected, the Gaussian controller performs better (has smaller attenuation). In the presence of non-Gaussian noise, however, the H^∞ controller performs sometimes better and sometimes worse (note that we did not try to check the behavior under the worst-case noise, which is typically linear in the state). Indeed, in simulations 1 and 2, we see that for the sinusoidal perturbation, the attenuation is better under an H^∞ controller, whereas in simulations 3 and 4, the situation is reversed.

8 Concluding Remarks

We presented in this paper a control theoretic approach for designing optimal rate-based flow control for high speed networks. Using the dynamic-game H^∞ approach that was developed in the last decade, we were able to handle very general measurement noise and interfering traffic: these may be bursty, non-stationary, non ergodic, highly correlated and even periodic. In spite of this generality in the description of the noise and interfering traffic, the resulting (optimal) controller is easy to implement, being linear in the (augmented) state of the system.

It turned out that the underlying control problem did not fall precisely into the standard H^∞ setting with imperfect information, and to bring it to the standard setting, the perfect state measurements had to be perturbed by small noise. The limit of the solution thus obtained, as the perturbation noise vanishes, provided the optimal control for the original problem.

We further studied the problem of delayed information; due to the Certainty-Equivalence property of both H^∞ and LQG optimal control (see [4]), it turned out that the problem with delayed information did not require an increase in the dimension of the state space (as is generally the case in optimal control problems with delayed information, see e.g. [3]). The optimal controller thus remains simple, and easy to implement also in the presence of delay, and thus appropriate for applications in high speed networks.

References

[1] The ATM Forum Technical Committee, *Traffic Management Specification*, Version 4.0, ATM Forum/95-0013R8, October 1995.

[2] E. Altman, F. Baccelli and J. C. Bolot, "Discrete-time analysis of adaptive rate control mechanisms", *High Speed Networks and their performance*, H. G. Perros and Y. Viniotis Eds., North Holland, pp. 121-140, 1994 (Proceedings of the 5th International Conference on High Speed Networks, Raleigh, USA, October 1993).

[3] E. Altman and P. Nain, "Closed-loop control with delayed information", *Proceedings of the ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, (New Port, Rhode Island), pp. 193–204, ACM, June 1992.

[4] T. Başar and P. Bernhard, *H^∞-Optimal Control and Relaxed Minimax Design Problems: A Dynamic Game Approach*, Birkhäuser, Boston, MA, 1991 (2nd edition, 1995).

[5] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Vols. 1 and 2, Athena Scientific, Belmont, Mass. 1995.

[6] J.-C. Bolot, "End-to-end delay and loss behavior in the Internet," *Proc. ACM Sigcomm '93*, pp. 289-298, San Fransisco, CA, Sept.1993.

[7] J-C. Bolot, "A self-tuning regulator for overload control in communication networks," *Proc. 31st Conf. on Decision and Control*, Tucson, AZ, pp. 1022-1023, Dec. 1992.

[8] J-C. Bolot and A. U. Shankar, "Analysis of a fluid approximation to flow control dynamics," *Proc. IEEE INFOCOM '92*, Florence, Italy, May 1992, pp. 2398-2407.

[9] F. Bonomi and K. W. Fendick, "The rate-base flow control framework for the Available Bit Rate ATM service", *IEEE Network*, pp. 25-39, March/April 1995.

[10] G. Didinsky, Z. Pan and T. Başar, "Parameter identification for uncertain plants using H^∞ methods," *Automatica*, vol. 31, no. 9, pp. 1227-1250, September 1995.

[11] T. Faber, L. H. Landweber and A. Mukherjee, "Dynamic time windows: Packet admission control with feedback," *Proc. ACM SIGCOMM '92*, Baltimore MD, Aug. 1992, pp. 124-135.

[12] K. Fendick, M. A. Rodrigues and A. Weiss, "Analysis of a rate-based control strategy with delayed feedback," *Proc. ACM SIGCOMM '92*, Baltimore, MD, August 1992, pp. 136-148.

[13] V. Jacobson, "Congestion avoidance and control," *Proc. ACM SIGCOMM '88*, Stanford, CA, pp. 314-329, August 1988.

[14] S. Keshav, "A control-theoretic approach to flow control," *Proc. ACM SIGCOMM '91*, Zurich, Switzerland, September 1991.

[15] H. T. Kung and R. Morris, "Credit-based flow control for ATM networks", *IEEE Network*, pp. 40-48, March/April 1995.

[16] K. K. Ramakrishnan and P. Newman, "Integration of rate and credit schemes for ATM flow control", *IEEE Network*, pp. 49-56, March/April 1995.

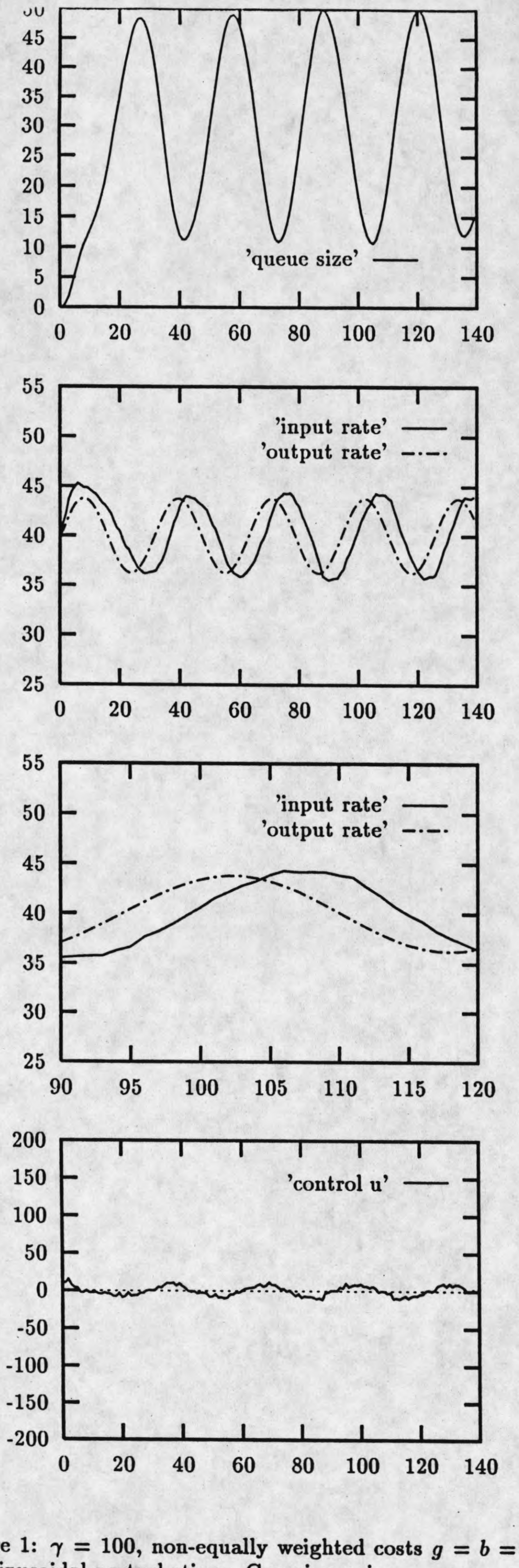

Figure 1: $\gamma = 100$, non-equally weighted costs $g = b = 0.1$, sinusoidal perturbations, Gaussian noise.

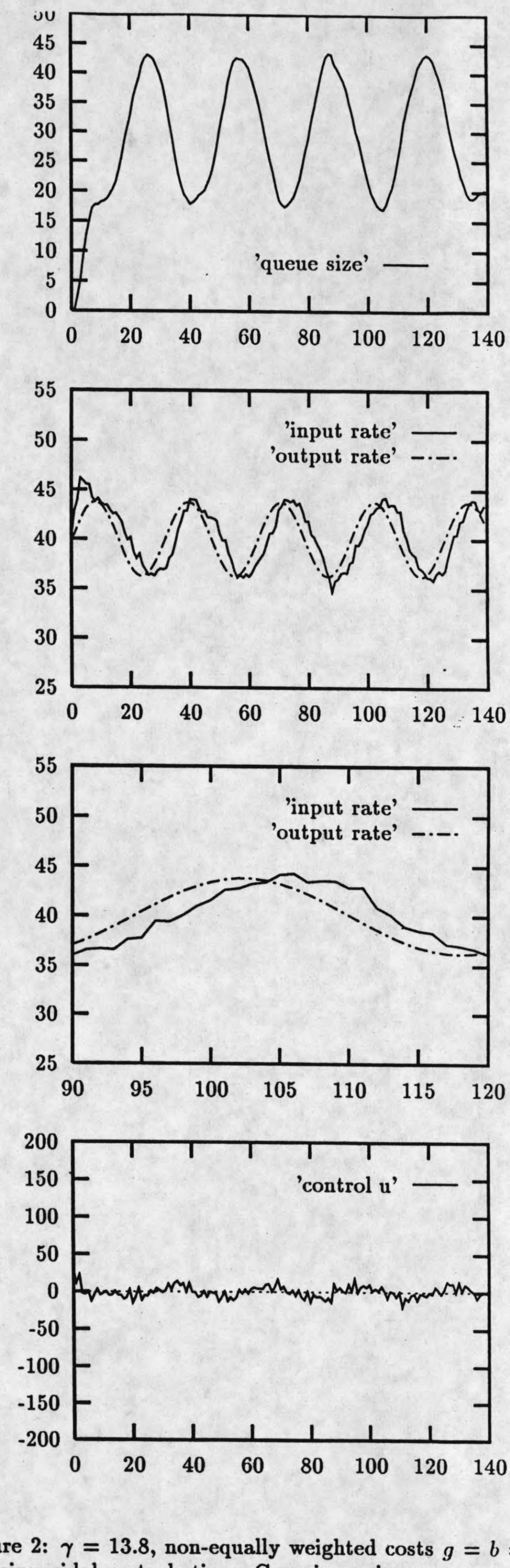

Figure 2: $\gamma = 13.8$, non-equally weighted costs $g = b = 0.1$, sinusoidal perturbations, Gaussian noise.

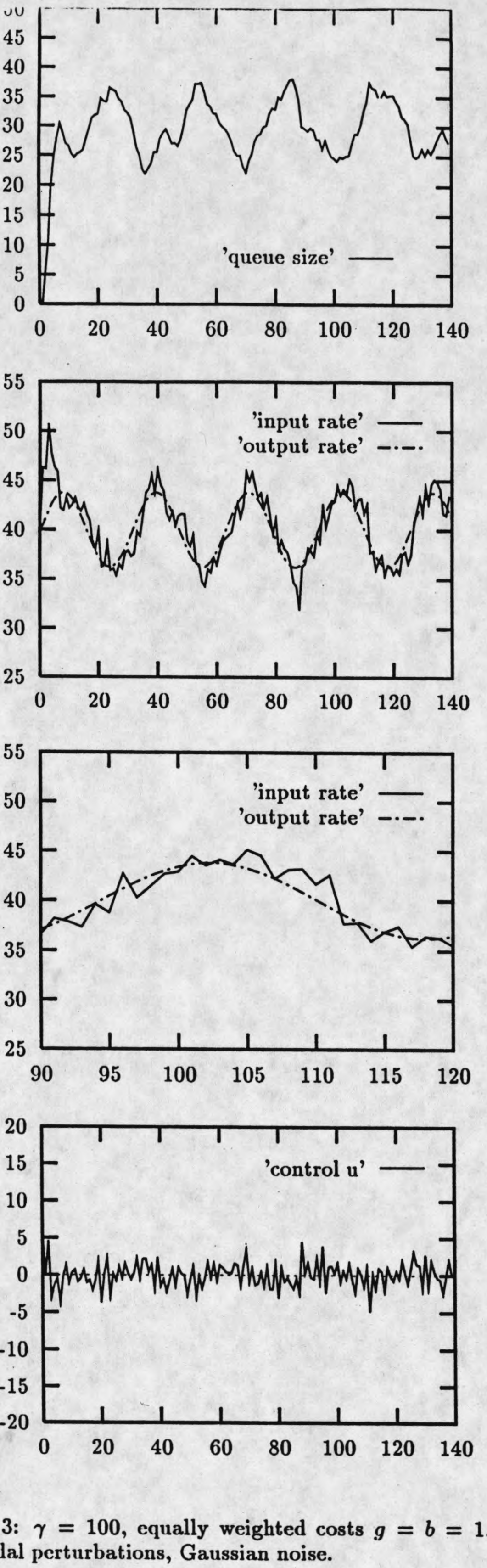

Figure 3: $\gamma = 100$, equally weighted costs $g = b = 1$, sinusoidal perturbations, Gaussian noise.

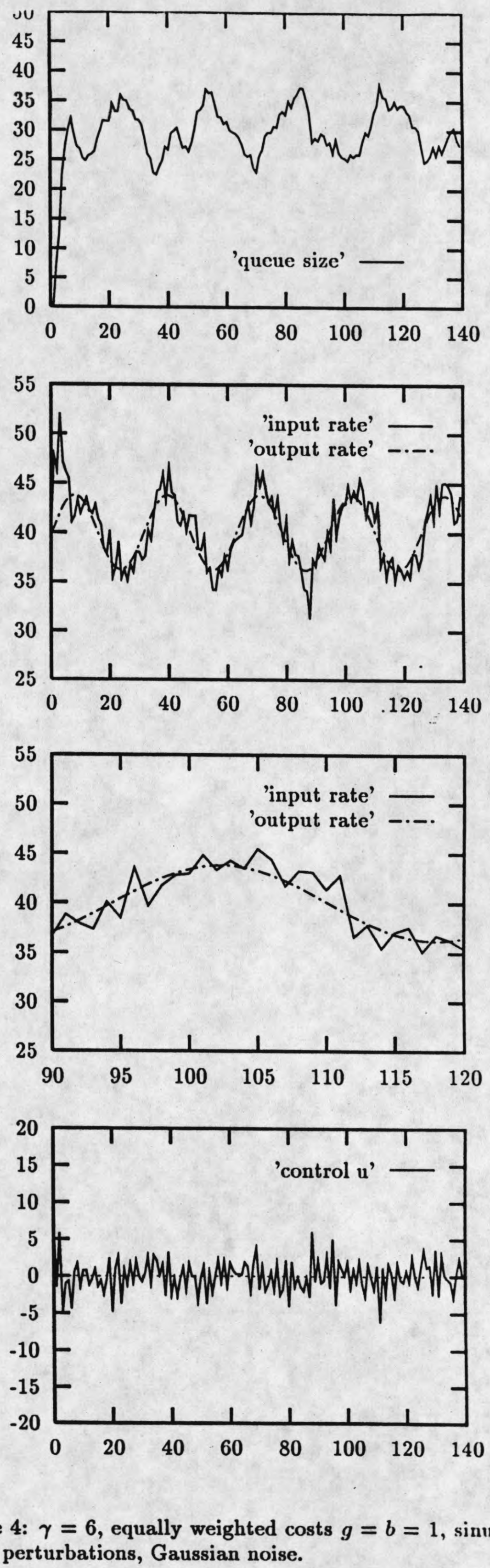

Figure 4: $\gamma = 6$, equally weighted costs $g = b = 1$, sinusoidal perturbations, Gaussian noise.

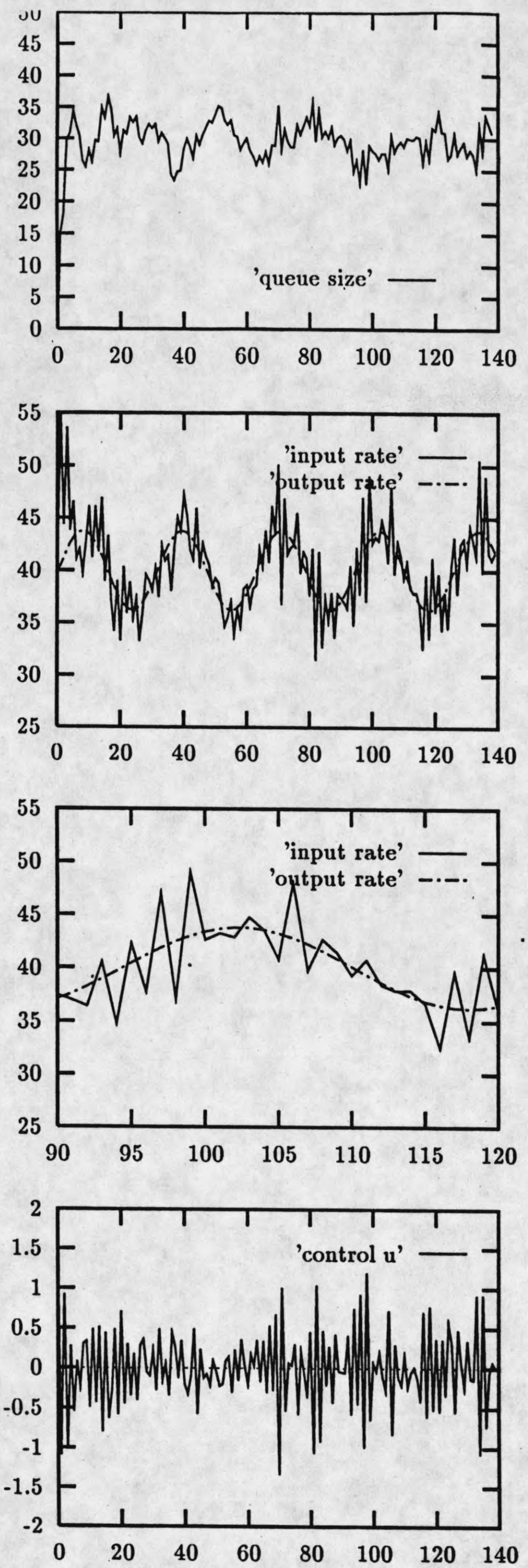

Figure 5: $\gamma = 100$, unequally weighted costs $g = 0.1, b = 10$, sinusoidal perturbations, Gaussian noise.

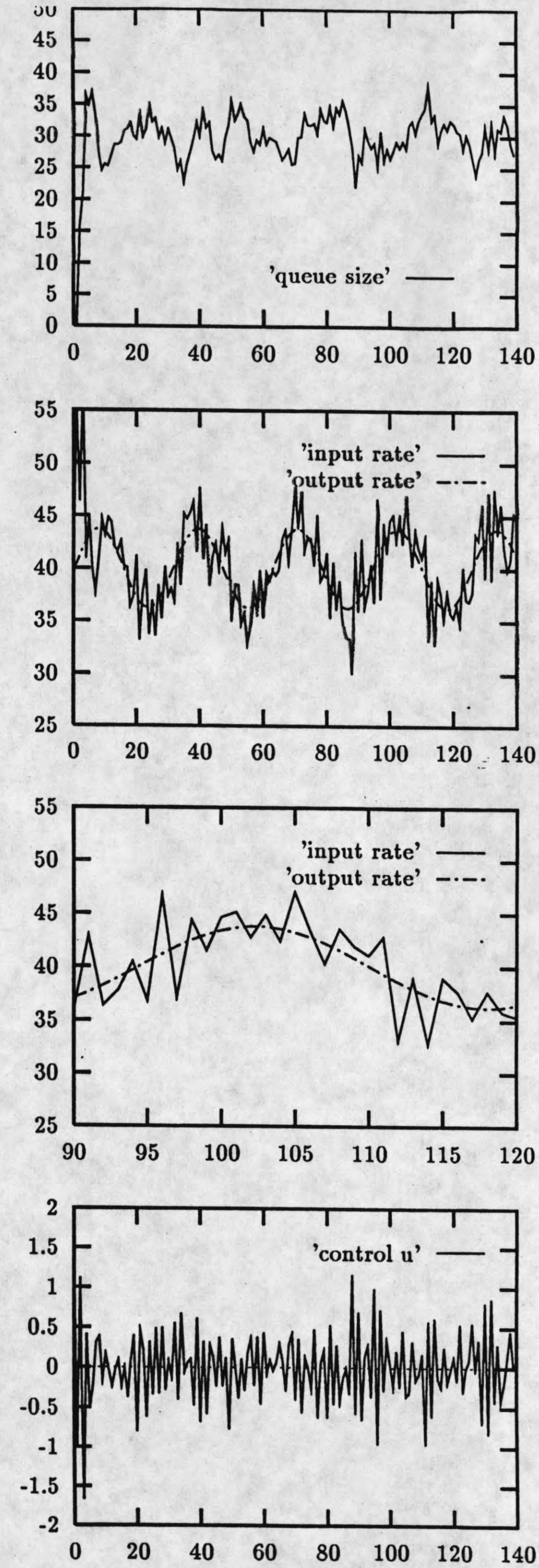

Figure 6: $\gamma = 3.8$, unequally weighted costs $g = 0.1, b = 10$, sinusoidal perturbations, Gaussian noise.

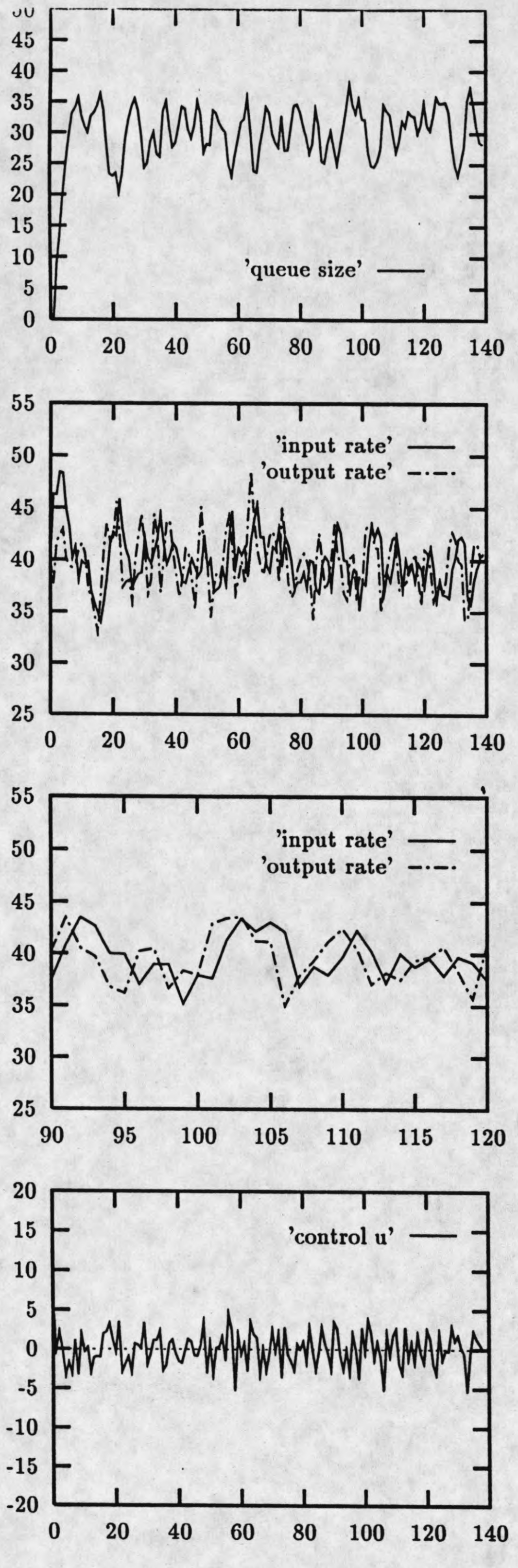

Figure 7: $\gamma = 100$, equally weighted costs $g = b = 1$, Gaussian noise and perturbations.

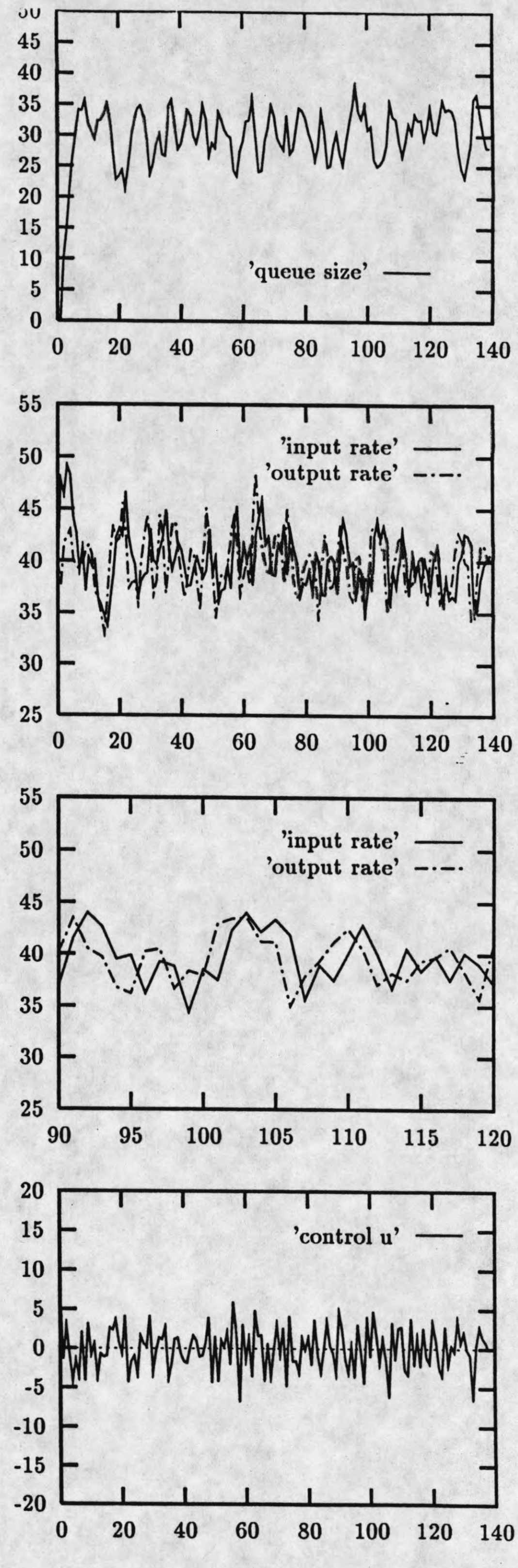

Figure 8: $\gamma = 6$, equally weighted costs $g = b = 1$, Gaussian noise and perturbations.

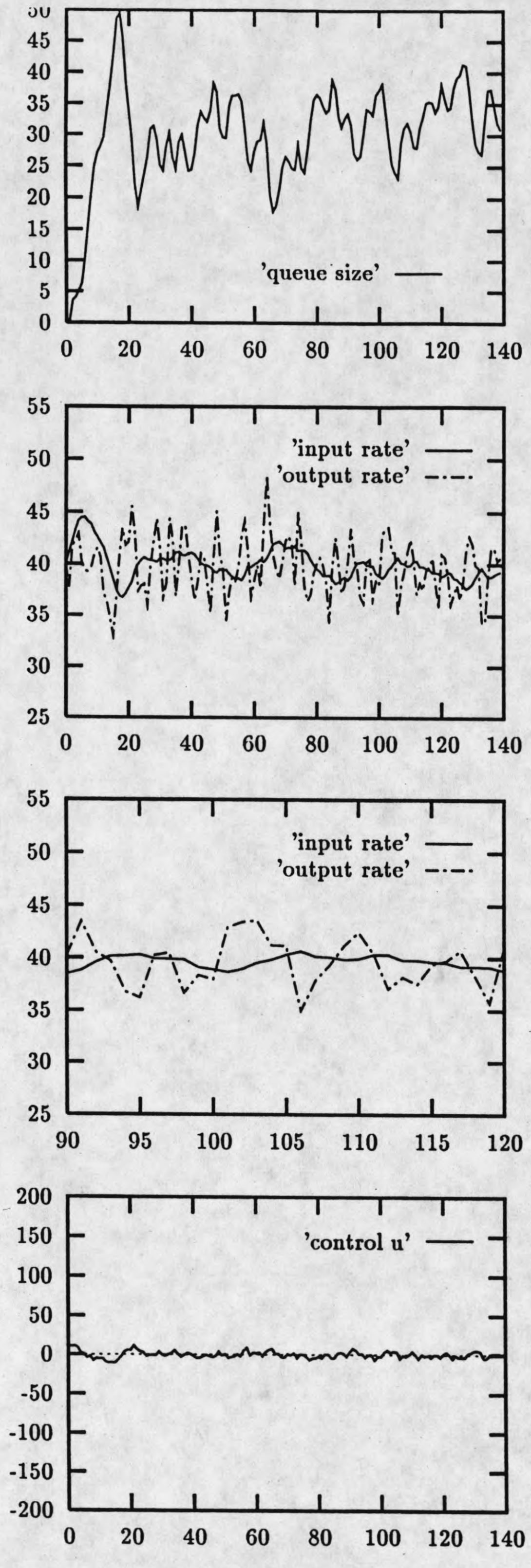

Figure 9: $\gamma = 100$, unequally weighted costs $g = b = 0.1$, Gaussian noise and perturbations.

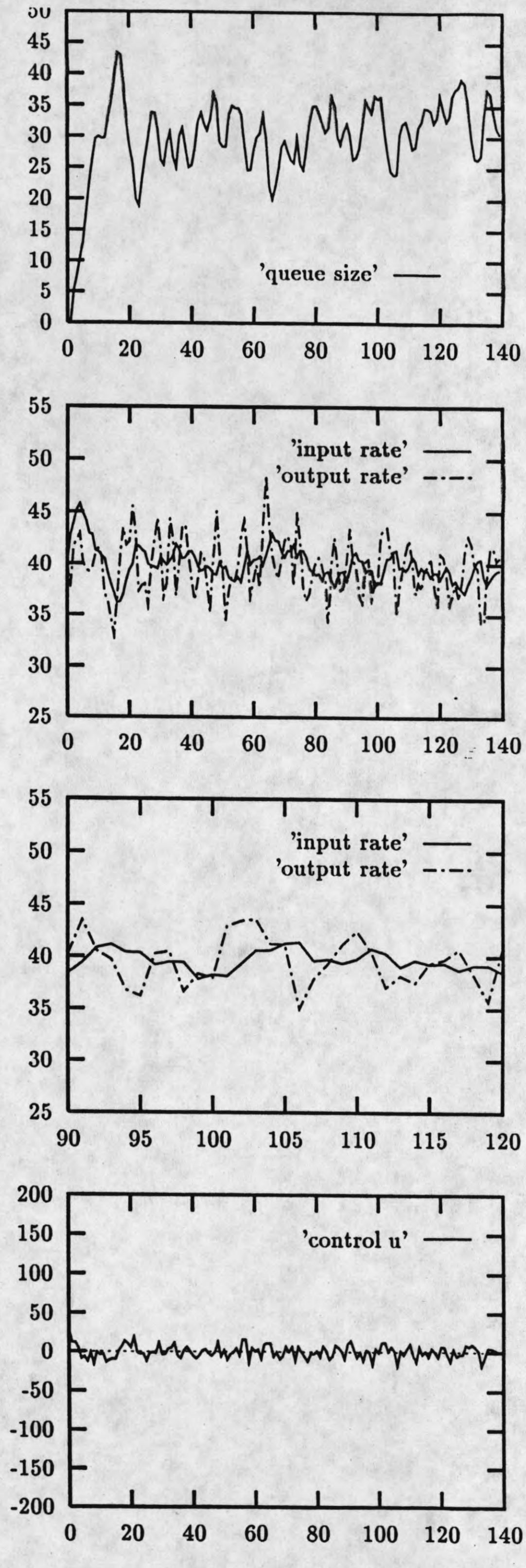

Figure 10: $\gamma = 13.8$, unequally weighted costs $g = b = 0.1$, Gaussian noise and perturbations.

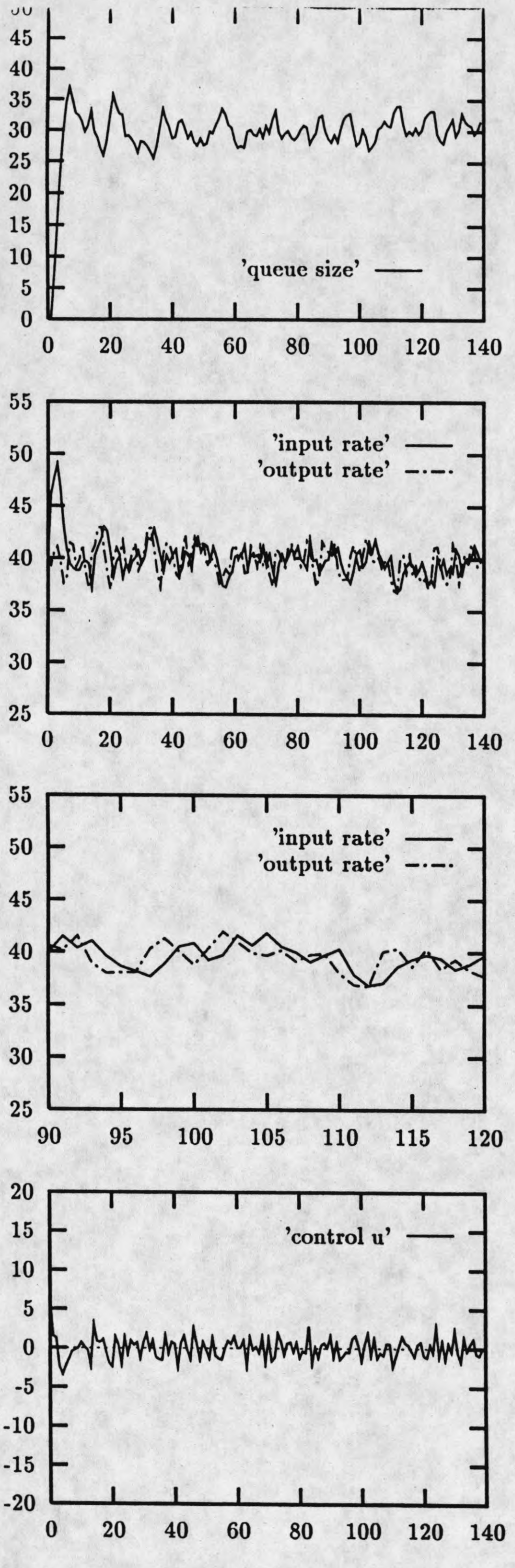

Figure 11: $\gamma = 100$, equally weighted costs $g = b = 1$, Uniformly distributed noise and perturbations.

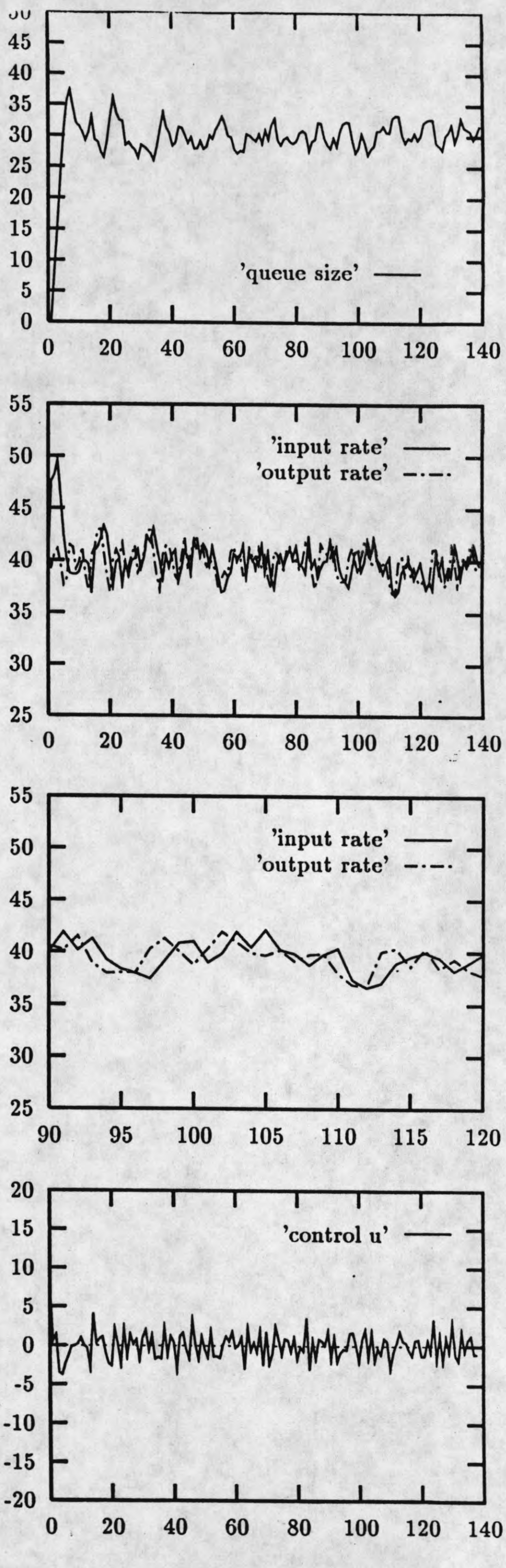

Figure 12: $\gamma = 6$, equally weighted costs $g = b = 1$, Uniformly distributed noise and perturbations.

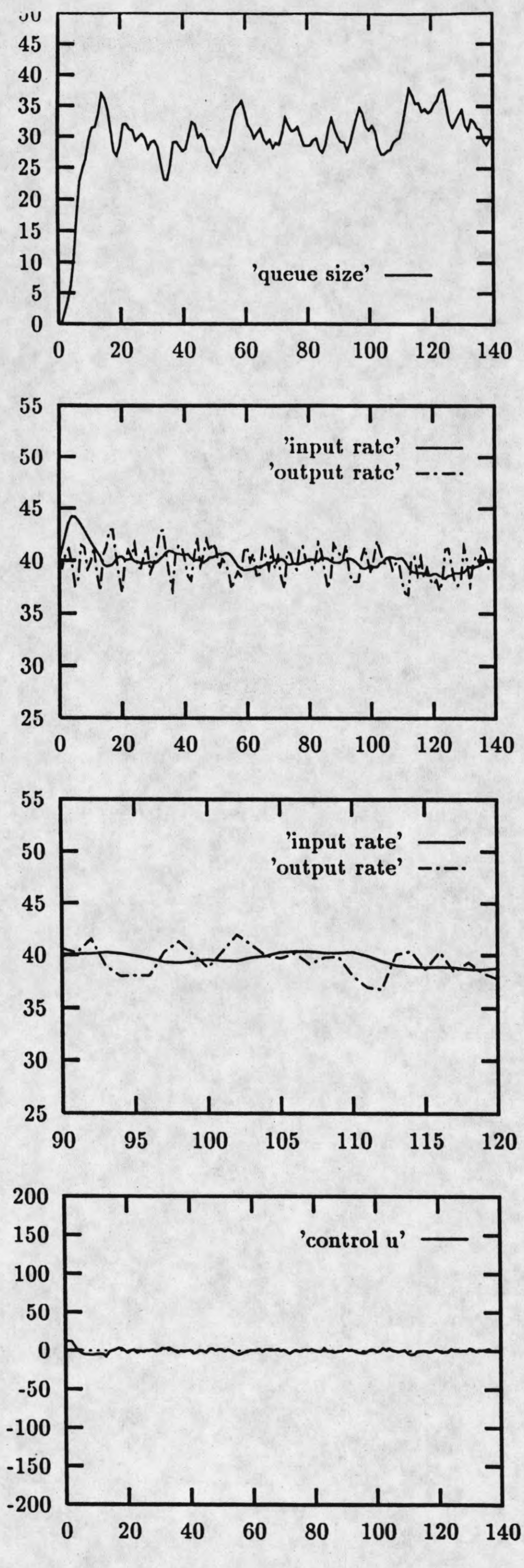

Figure 13: $\gamma = 100$, unequally weighted costs $g = b = 0.1$, Uniformly distributed noise and perturbations.

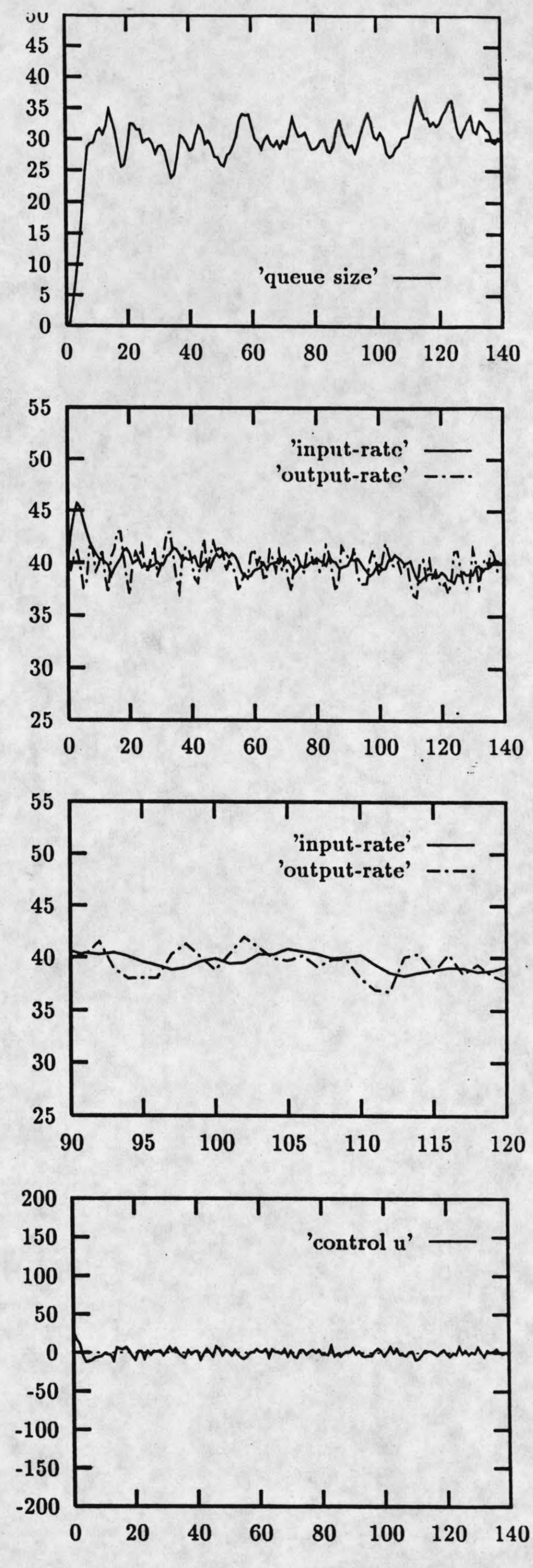

Figure 14: $\gamma = 13.8$, unequally weighted costs $g = b = 0.1$, Uniformly distributed noise and perturbations.