Conundrums in everyday stochastic control

The excitation price of performance

Bob Bitmead
Mech & Aero Engineering, UCSD

Proto-Conclusions

Agile stochastic control systems without full state information require the control signal to provide persistent probing to facilitate state estimation

This comes at a cost to the estimation-free control performance

The control signal balances regulation and excitation - *duality*

It is inherently part of stochastic optimal control and is very hard to optimize computationally

Three examples of commercial suboptimal stochastic control will be presented - wireless power control, TCP/IP, hepatitis B management

Each contains dual action

expending energy to save energy

consuming capacity to save capacity

spending money to save money

You pay to use them

Engineered Agile Stochastic Control Systems

Objectives

- Part 1 stochastic optimal control
  - background ideas
    - duality and probing are inherent
- Part 2 familiar agile control system in communications
  - power control in mobile telephony
    - suboptimal practical solution
    - attempt at optimality
- Part 3 familiar agile control in networks
  - internet congestion control
    - testing the constraints
- Part 4 duality in healthcare
  - paying to save money
- Part 5 robotics and autonomy
  - enforcing observability in hostile environments
- Some outrageous comments about the future
  - good luck with that

Part 1: stochastic optimal control background

It is a demanding business

Stochastic optimal control problem

System — **Markov process**

$$x_{k+1} = f_k(x_k, u_k, w_k)$$
$$y_k = h_k(x_k, u_k, v_k)$$

noises

starting information $\pi_0 = p(x_0)$ the initial state density

objective $\text{minimize } J_N = \mathrm{E}\left[\sum_{\ell=0}^{N-1} c_\ell(x_{k+\ell}, u_{k+\ell}) + c_N(x_{k+N})\right]$

stage cost

terminal cost

expectation is taken over distribution of $\{\pi_0, v_0, w_0, v_1, \ldots, v_{N-1}, w_{N-1}\}$

constraints

subject to: $u_\ell \in \mathcal{U}_\ell$ — **control constraints**

$x_\ell \in \mathcal{X}_\ell$ — **state constraints**

$x_N \in \mathcal{X}_N$ — **terminal constraint**

admissible controls $u_\ell = g_\ell(Z^\ell)$ where $Z^\ell = \{y_\ell, u_{\ell-1}, y_{\ell-1}, u_{\ell-2}, \ldots, y_0\}$

nonanticipative control

The solution depends solely on the *Information State*: $\pi_\ell = p\left(x_\ell | Z^\ell\right)$

i.e. $u_\ell = g_\ell(\pi_\ell)$

a "separated" control

Stochastic optimal control solution

The *Bayesian filter* propagates the information state using $u_{\ell-1},\ y_\ell$

$$\pi_{\ell-1} = p(x_{\ell-1}|Z^{\ell-1}) \to \pi_\ell = p(x_\ell|Z^\ell)$$

from state equation

$$\pi_\ell^- = p(x_\ell|Z^{\ell-1}) = \int_{x_{\ell-1}} p(x_\ell|x_{\ell-1}, u_{\ell-1})\pi_{\ell-1}\, dx_{\ell-1}$$

time update

$$\pi_\ell = p(x_\ell|Z^\ell) = p(y_\ell|x_\ell)\pi_\ell^- \times \frac{1}{\int_{x_\ell} p(y_\ell|x_\ell)\pi_\ell^-\, dx_\ell}$$

measurement update

from measurement equation

Call this $\pi_\ell = T_{\ell-1}[\pi_{\ell-1}, y_\ell, u_{\ell-1}]$

Bayesian filter state conditional density update; depends on $u_{\ell-1}, y_\ell$

The *Stochastic Dynamic Programming Equation* yields the control u_ℓ and value

$$V_\ell(\pi_\ell) = \min_{u_\ell \in \mathcal{U}_\ell} \left\{ \int_{x_\ell} c_\ell(x_\ell, u_\ell)\pi_\ell\, dx_\ell + \int_{y_{\ell+1}} V_{\ell+1}\left(T_\ell(\pi_\ell, y_{\ell+1}, u_\ell)\right) p(y_{\ell+1}|\pi_\ell, u_\ell)\, dy_{\ell+1} \right\}$$

starting from

$$V_N(\pi_N) = \int_{x_N} c_N(x_N) T_{N-1}[\ldots T_{\ell+1}[T_\ell, y_{\ell+1}, u_\ell], y_{\ell+2}, u_{\ell+1}] \ldots, y_N, u_{N-1}]\, dx_N$$

The information state recursion integrates densities **HARD!**

Stochastic Dynamic Programming integrates propagated densities **VERY HARD!**

and this is discrete time!

$$\min_{u_{N-3}} \int_{\Omega(y_{N-2})} A_{N-2}\, dy_{N-2} = \min_{u_{N-3}} \Bigg\{ \int_{\Omega(y_{N-2})} \sum_i \left(\frac{f_i^2 u_{N-3}}{\sigma_w^2} - \gamma^*\right)^2 \frac{\pi_{N-3}(i)}{\sqrt{2\pi}\,\sigma_w} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-2} - f_i\sqrt{u_{N-3}}\right)^2\right] dy_{N-2}$$

$$+ \int_{\Omega(y_{N-2})} \min_{u_{N-2}} \Bigg\{ \int_{\Omega(y_{N-1})} \sum_i \left(\frac{f_i^2 u_{N-2}}{\sigma_w^2} - \gamma^*\right)^2 \frac{\pi_{N-3}(i)}{\sqrt{2\pi}\,\sigma_w M_{N-2}} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-2} - f_i\sqrt{u_{N-3}}\right)^2\right]$$

$$\times \frac{1}{\sqrt{2\pi}\,\sigma_w} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-1} - f_i\sqrt{u_{N-2}}\right)^2\right] dy_{N-1}$$

$$+ \int_{\Omega(y_{N-1})} \min_{u_{N-1}} \Bigg\{ \int_{\Omega(y_N)} \sum_i \left(\frac{f_i^2 u_{N-1}}{\sigma_w^2} - \gamma^*\right)^2 \frac{\pi_{N-3}(i)}{\sqrt{2\pi}\,\sigma_w M_{N-2}} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-2} - f_i\sqrt{u_{N-3}}\right)^2\right]$$

$$\times \frac{1}{\sqrt{2\pi}\,\sigma_w M_{N-1}} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-1} - f_i\sqrt{u_{N-2}}\right)^2\right]$$

$$\times \frac{1}{\sqrt{2\pi}\,\sigma_w} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_N - f_i\sqrt{u_{N-1}}\right)^2\right] dy_N \Bigg\} \to \min_{u_{N-1}}$$

$$\times \sum_i \frac{\pi_{N-3}(i)}{\sqrt{2\pi}\,\sigma_w M_{N-2}} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-2} - f_i\sqrt{u_{N-3}}\right)^2\right]$$

$$\times \frac{1}{\sqrt{2\pi}\,\sigma_w} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-1} - f_i\sqrt{u_{N-2}}\right)^2\right] dy_{N-1} \Bigg\} \to \min_{u_{N-2}}$$

$$\times \sum_i \frac{\pi_{N-3}(i)}{\sqrt{2\pi}\,\sigma_w} \exp\left[-\frac{1}{2\sigma_w^2}\left(y_{N-2} - f_i\sqrt{u_{N-3}}\right)^2\right] dy_{N-2} \Bigg\} \to \min_{u_{N-3}}$$

Stochastic optimal control

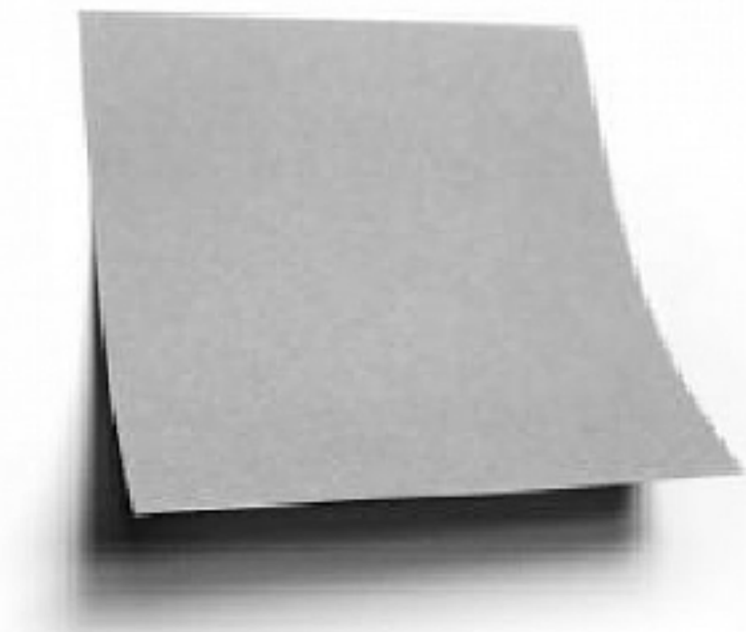
1. The solution process comprises a coupled recursion:
   - information state, value function
2. Duality/learning/probing are inherent in the optimality
   - The dependence of the future information state on the current control is incorporated via the T_k operator
   - If probing improves the value then it is included via the recursion
3. All feasible future control sequences and information states are explored and averaged in SDPE
   - There is no concept of a future optimal trajectory
   - There is no concept of a future sequence of information states
     - SDPE acts as the oracle

       Give it the current information state density π_ℓ

       it returns the optimal control value for the current step

       It implements the optimal feedback policy

       but it does not reveal it
4. It is the optimal control



Message 1

Stochastic optimal control provides a framework for output (state-estimate) feedback which inherently accommodates probing and learning

Probing costs control effort but is part of the optimal control
You cannot do better with less excitation

Part 2: familiar learning systems

engineered adaptive control systems which work without human intervention billions of times per second

Curating Defunct Non-Agile Technologies

What is going on here? We know these signals so well!!

- They are training signals known to both ends
  - No information content but used to *identify* the channel
    - In the case of faxes and modems this is used to set data rates for reliable communication
      - A once-off transaction in a network of connections
      - A small overhead and transaction cost

What about packet-switched mobile networks

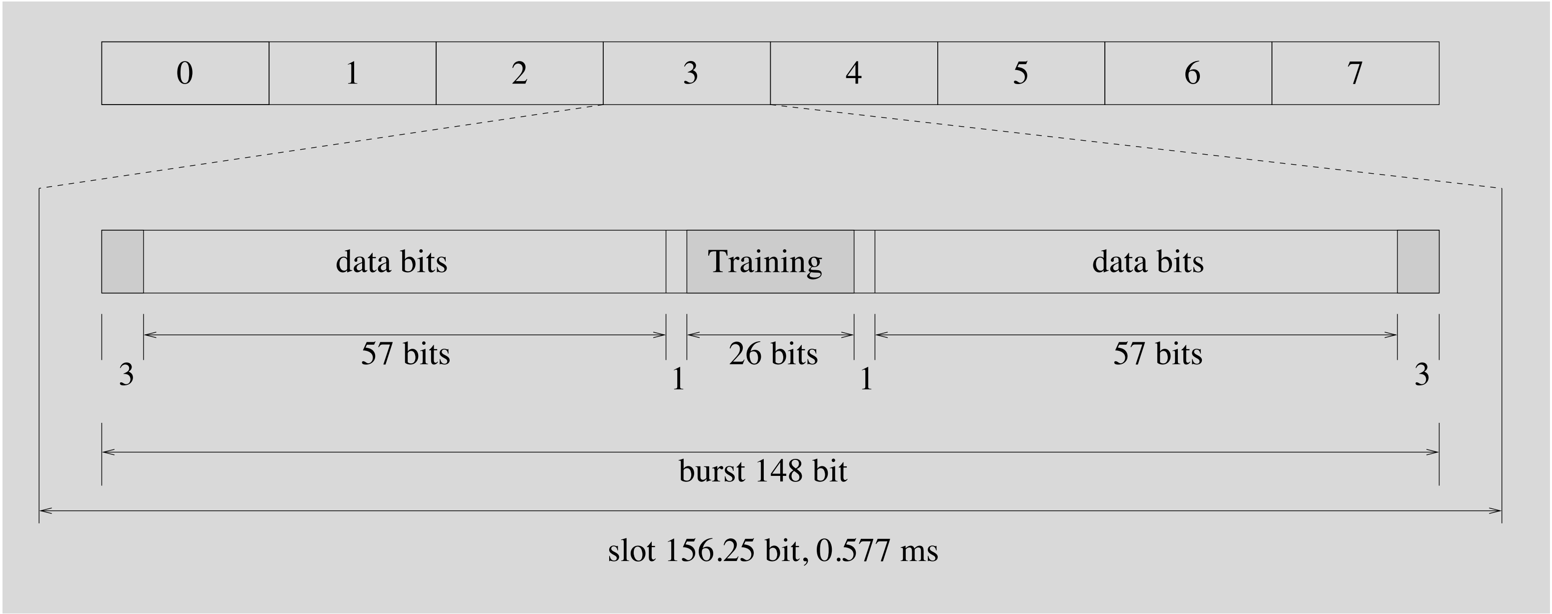
- The channel changes on a 10s of milliseconds basis
  - The price of connection is eternal vigilance

Agile Channel Modeling in Mobile Wireless

Mobility leads to rapid channel variation

- Reflections plus fading
  - We need power control
  - We need equalization
    - We need a channel model!

The GSM mid-amble:



<u>Every</u> packet: 26 bits of training per 114 bits of data

Base station computes a channel model - equalization of data
determines mobile station SNR at receiver - power control

Opportunity cost of channel modeling

Roughly 19% of all channel capacity is turned over to transmission of information-free training signals

- Without this the system does not function
  - The channel model is a 6-tap FIR filter

What about power control - one parameter: fade

- Why power control?
  - battery life
  - interference with other users
    - This is an energy management problem
- The channel fade is dependent on distance from tower and path
  - This changes rapidly with mobility
    - The base station estimates SNR and instructs the mobile to change power by ±2 dB every packet

How much energy does it take to save energy?

The fundamental conundrum

Additive white gaussian noise channel: $z_k = f u_k + w_k$

- Fade f, signal u_k ,noise $w_k \sim \mathcal{N}(0, \sigma_w^2)$
  - Bayes rule for updating the fade density

$$p(f|Z^k) = \frac{p(z_k|f)p(f|Z^{k-1})}{\int_f p(z_k|f)p(f|Z^{k-1})\,df}$$

new density → $p(f|Z^k)$; *prior density* → $p(f|Z^{k-1})$; *normalization* → $\int_f p(z_k|f)p(f|Z^{k-1})\,df$; *sharpening function* → $p(z_k|f)$

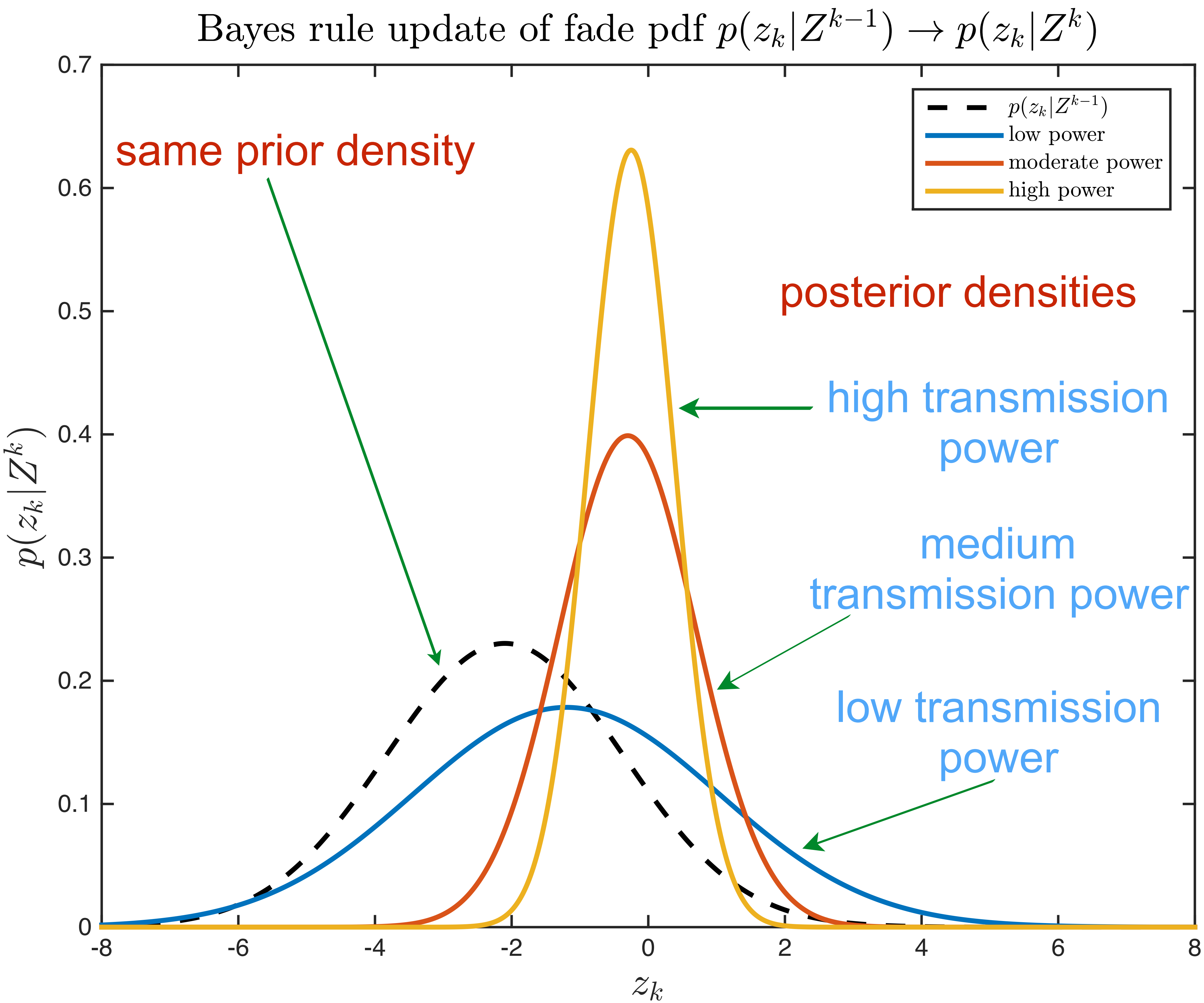
- in the gaussian case

$$p(z_k|f) = \frac{1}{2\sqrt{\pi}\sigma_w} \exp\left[-\frac{(z_k - f u_k)^2}{2\sigma_w^2}\right]$$

The more power, u_k^2, is used to send the training signal, the faster we learn the value of f

- You have to expend energy to save energy
  - This is sometimes called the exploration-exploitation trade-off
    - As a formal mathematical problem to optimize the energy, it goes back to Fel'dbaum in 1960
      - It is computationally intractable
      - The sticking point is optimality

Spending energy to save energy



different conditional means, different conditional variances

The opportunity cost in mobile power control

The system does not work without models

- Modeling requires energy via training signal power
  - Optimization of the energy budget is so tough as to cost even more energy in computation
    - You pay the phone company for a suboptimal system

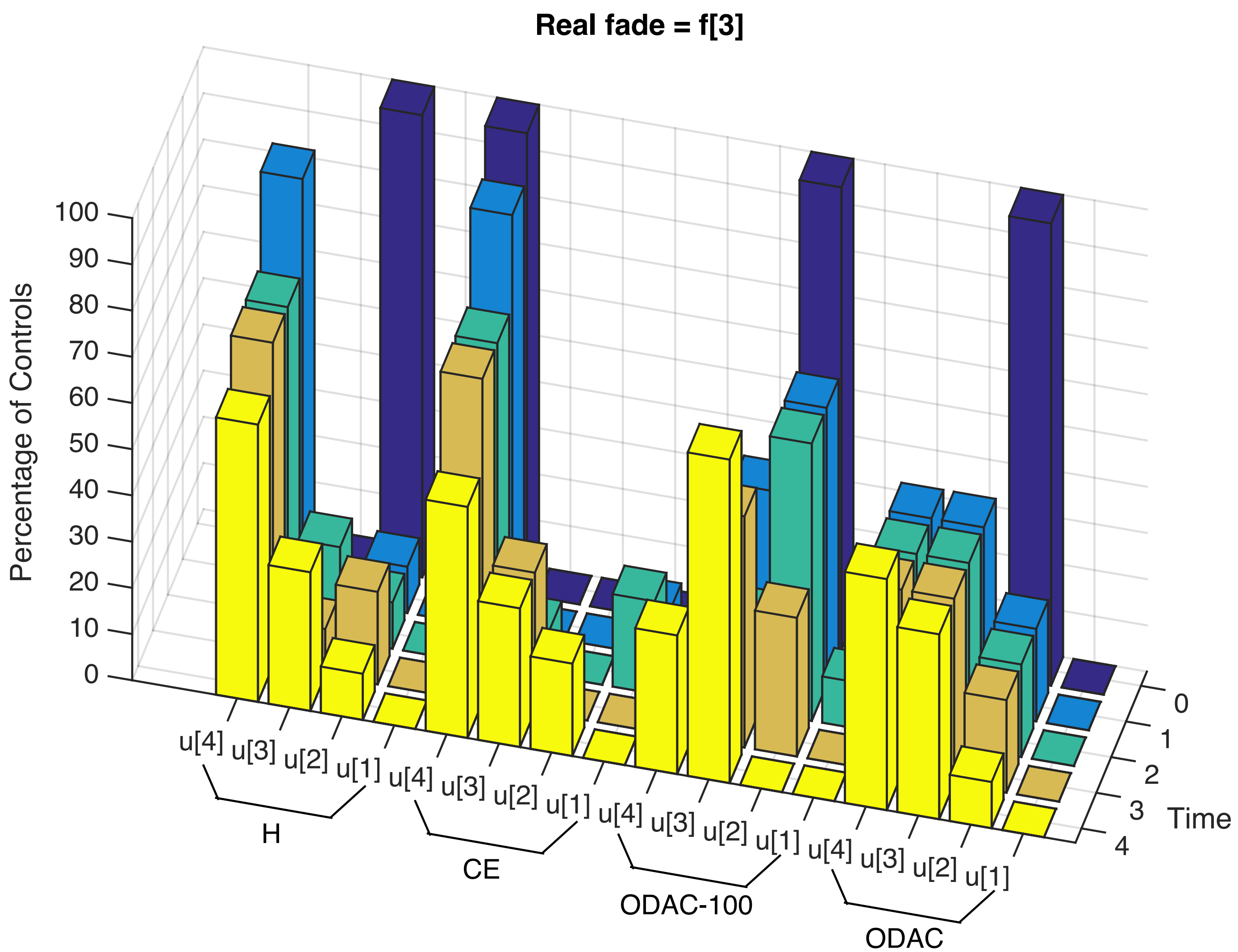


Fig. 3. Histograms of control values averaged over 10 realizations of the noise for each adaptive control scheme vs. time. The real fade is equal to $f[3] = -3$dB.

4 fade/power values
Gaussian noise
- zero mean
- known variance

5 steps

7 hours computation to yield optimal powers for a few microseconds

Message 2

Learning in agile systems is hard and comes with a cost

… and a definite benefit

Mobile telephony requires training signals in order to operate

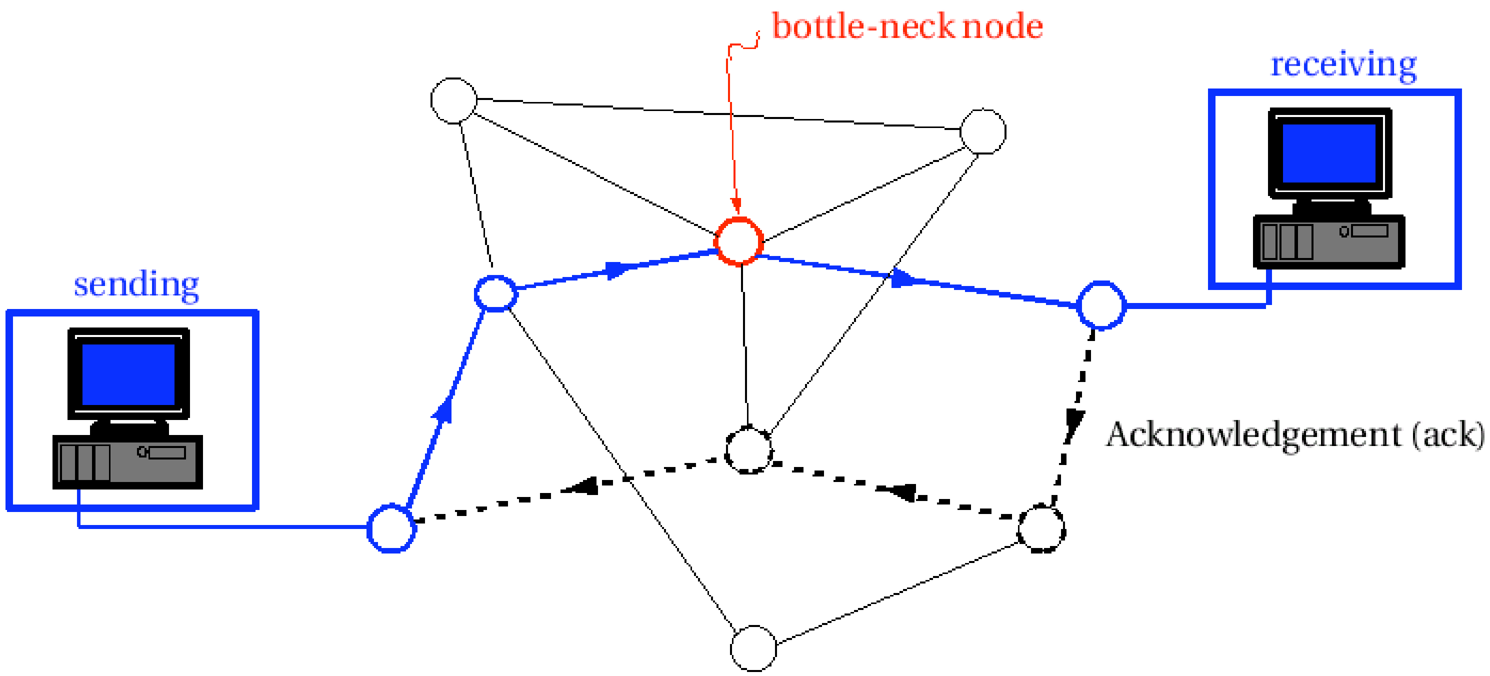
Optimal design, were it tractable, would require accurate probabilistic models

… so we settle for suboptimality

Part 3: network congestion control

TCP/IP AIMD RED
strategies for the internet

TCP as feedback

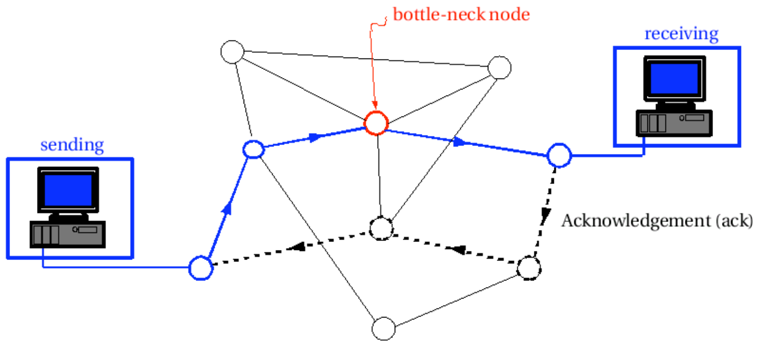


From the perspective of the source node

- packets are sent into the network
- a packet arriving at destination produces an ACK packet response
  - the send rate is controlled into network based on ACK sequence
    - the aim is to avoid traffic congestion
      - the competing traffic is stochastic in nature
- normally modeled as dominated by a single bottleneck node

This is a stochastic feedback control problem

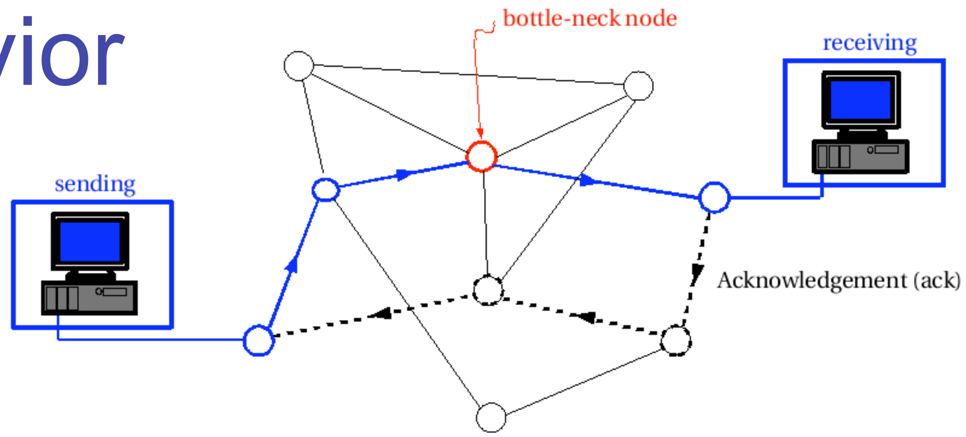
Destination node behavior



- Upon receipt of a packet from the source, the destination node sends an acknowledgement packet in reply
  - Generally ACKs are simple few-bit packets
    - arriving packet identifier
  - There are proposals for ACKs to contain more and more useful data
    - arrival time
    - data inserted by intervening nodes
      - buffer state and/or statistics
      - traffic state and/or statistics
  - This is no longer TCP/IP

- The ACKs have to travel back through the same network
  - They too can be lost

Source node behavior

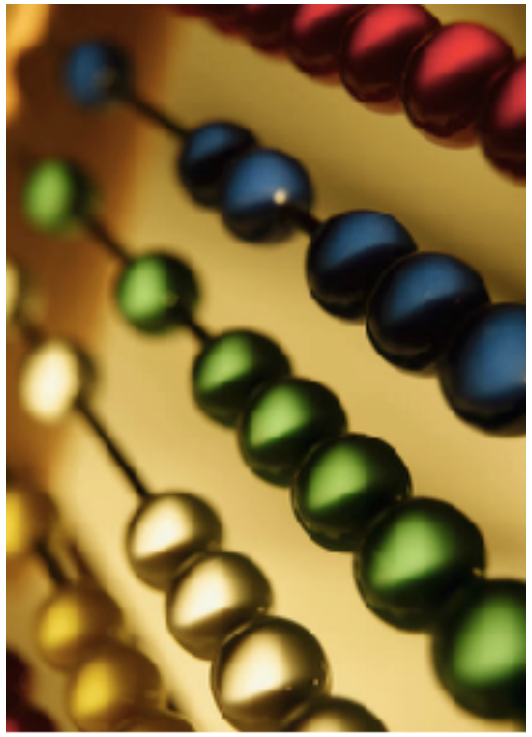


Data made up into packets with rate r_k

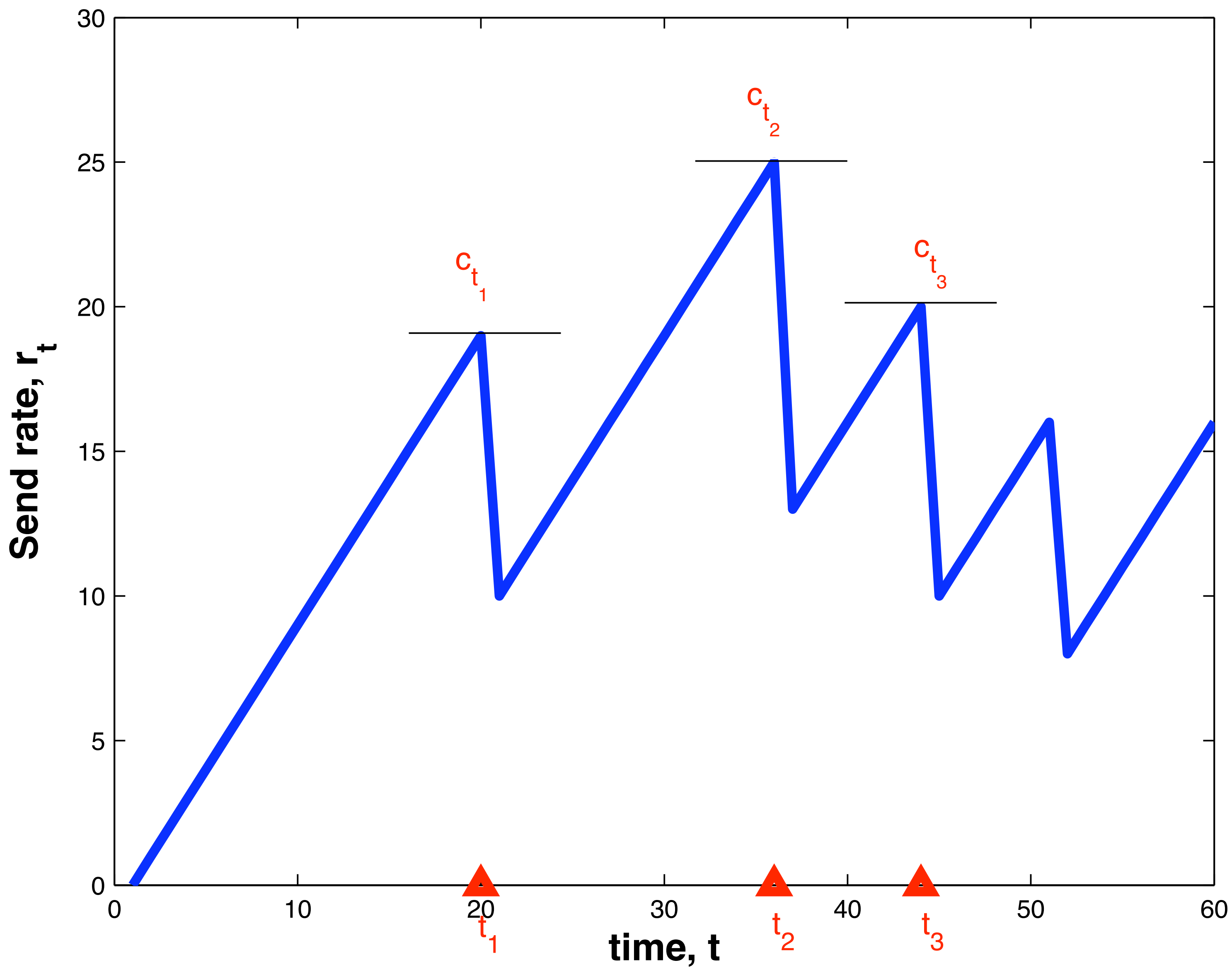
- Packet rate into network is the mechanism of congestion control
  - Rate r_k is adjusted in response to arrival or non-arrival of ACKs
    - non-arrival: time-out or ACK out of sequence

Common congestion control law AIMD

- Additive Increase / Multiplicative Decrease
  - ACK arrives: increase rate r_k by one packet
  - ACK missing: decrease r_k by a factor of two



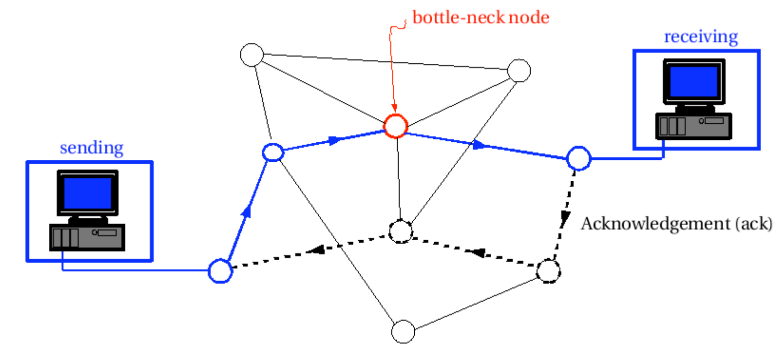
AIMD

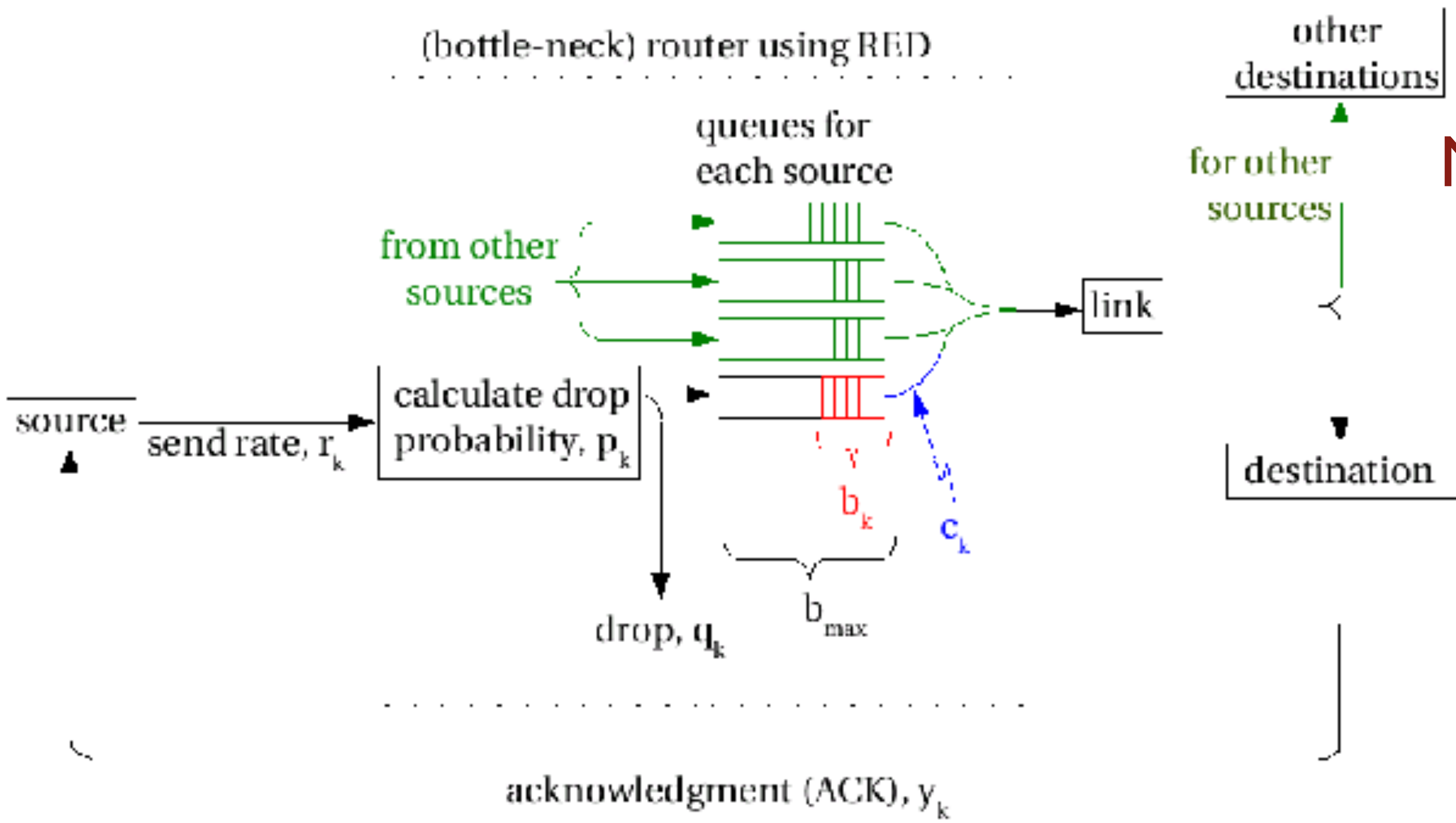


Additive increase / multiplicative decrease

- ACK arrives: increase rate r_k by one packet
- ACK missing: decrease r_k by a factor of two

Bottleneck node behavior





Nodes have finite buffers

buffer occupancy	b_k
competing traffic (random)	c_k
arrival rate from source	r_k
packets deliberately dropped	q_k

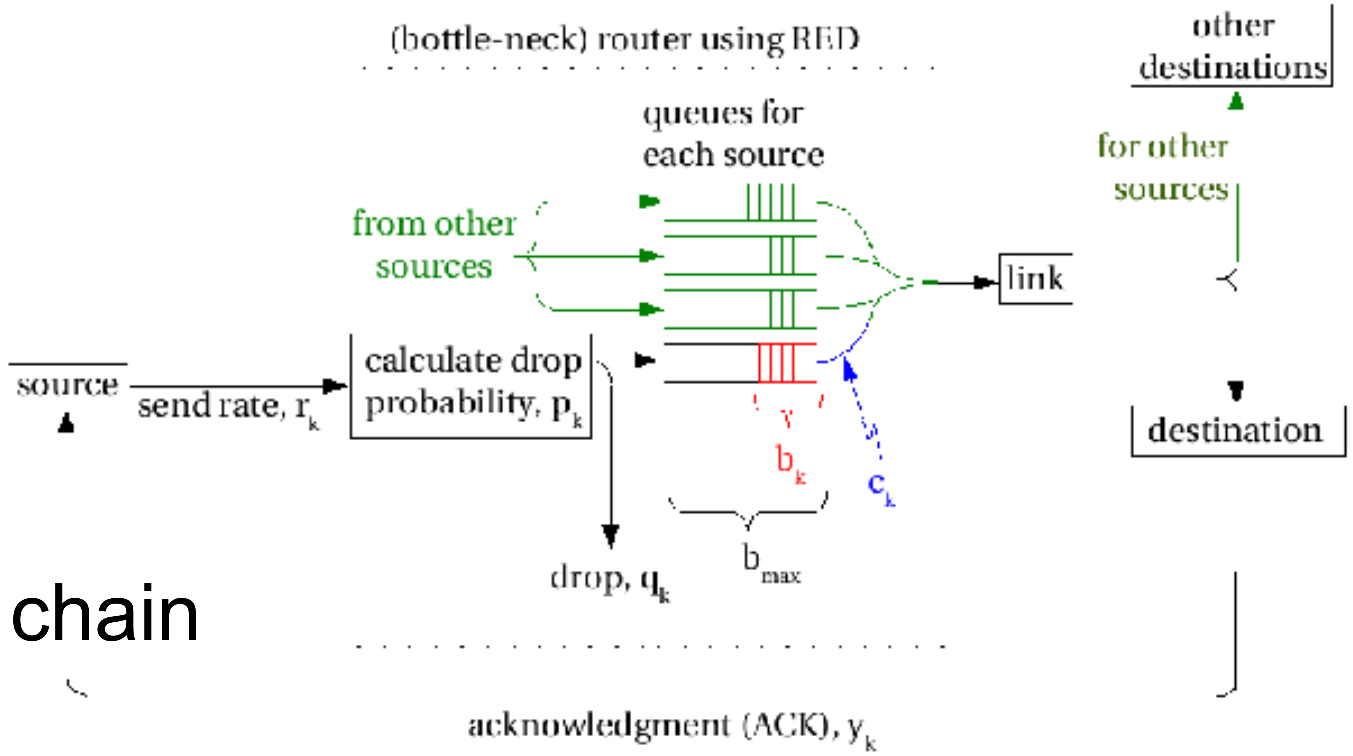
Packets are dropped with a certain probability p_k

- *Random Early Detection* (RED) algorithm
  - p_k depends on buffer state b_k

$$b_k \to \text{full} \implies p_k \to 1$$
$$b_k \to \text{empty} \implies p_k \to 0$$

- *Droptail* algorithm drops all packets only when buffer is full
- There are other algorithms

Hidden Markov Model of Bottleneck Node



State of the bottleneck node

$$x_k = \begin{pmatrix} b_k \\ c_k \end{pmatrix}$$

Drop probability $p_k = f(b_k)$

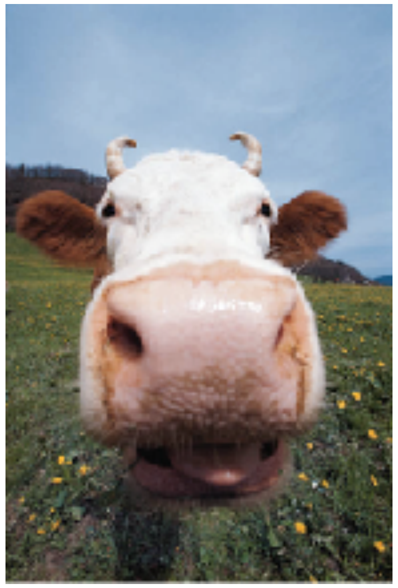
Capacity (traffic) model - Markov chain

$$P(c_{k+1} = c) = \sum_d P(c_{k+1} = c | c_k = d) P(c_k = d)$$

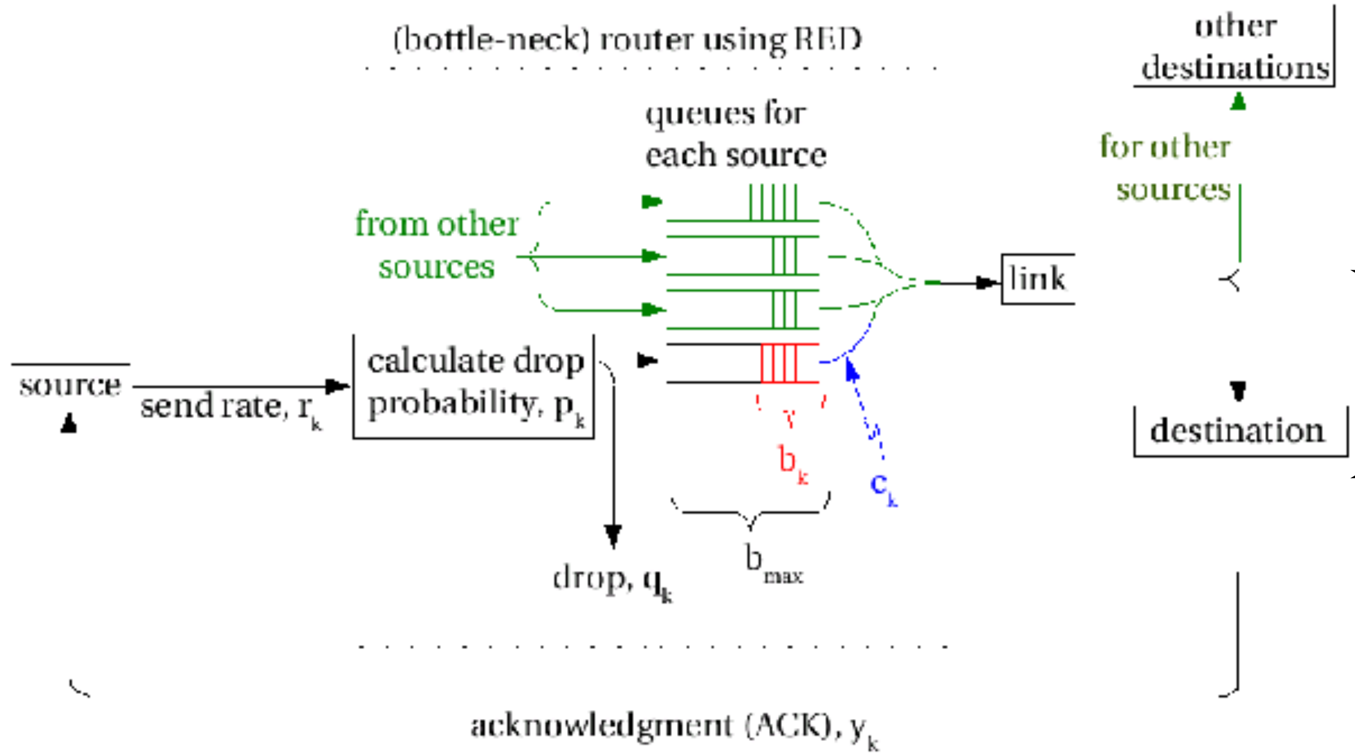
Hidden Markov Model for bottleneck node state

$$\begin{aligned} \Pi_x(k+1) &= A(r_k)\Pi_x(k) \\ \Pi_y(k) &= C\Pi_x(k) \end{aligned}$$

Known model - input sequence r_k and output (ACK) sequence y_k



A control theorist arrives on the scene



Can the source computer reliably estimate the state of the bottleneck?

- Buffer length and capacity value
  - Available data are input rate history and ACK sequence history

This is an observability question about an HMM

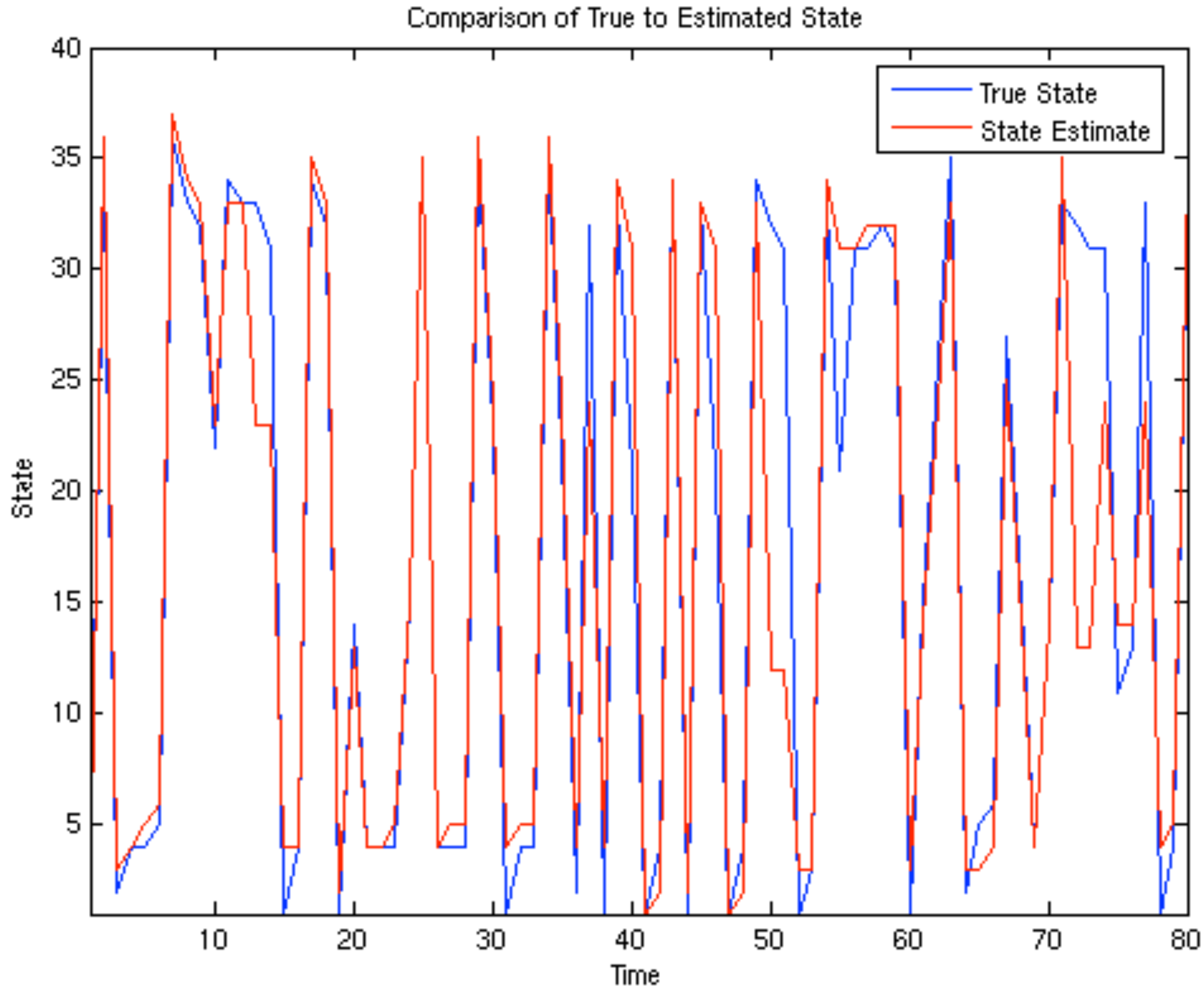
- A theory of HMM filtering and smoothing exists
  - But what about the observability questions?

Does the input-output sequence suffice to estimate the state (c_k, b_k)?

Yes with AIMD!

Performance of the HMM smoother

Capacity estimate



Message 3

TCP/AIMD/RED handle the observability of the varying capacity constraint by continually exceeding it … after which we need to retransmit the whole packet

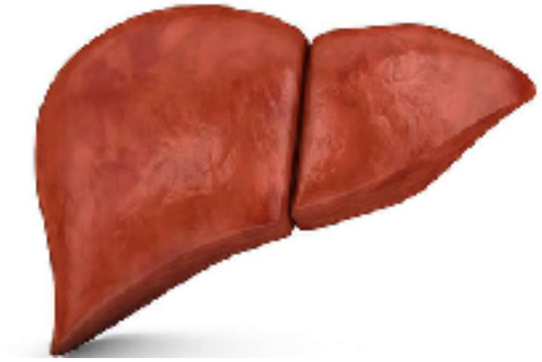
“Unless I exceed a constraint, how could I know its value?”

Bob Shorten, Dublin

Part 4: healthcare modeling and control

POMDP control and duality
Hidden Markov Modeling of Hepatitis B

hepatitis B progression model



hepatitis B

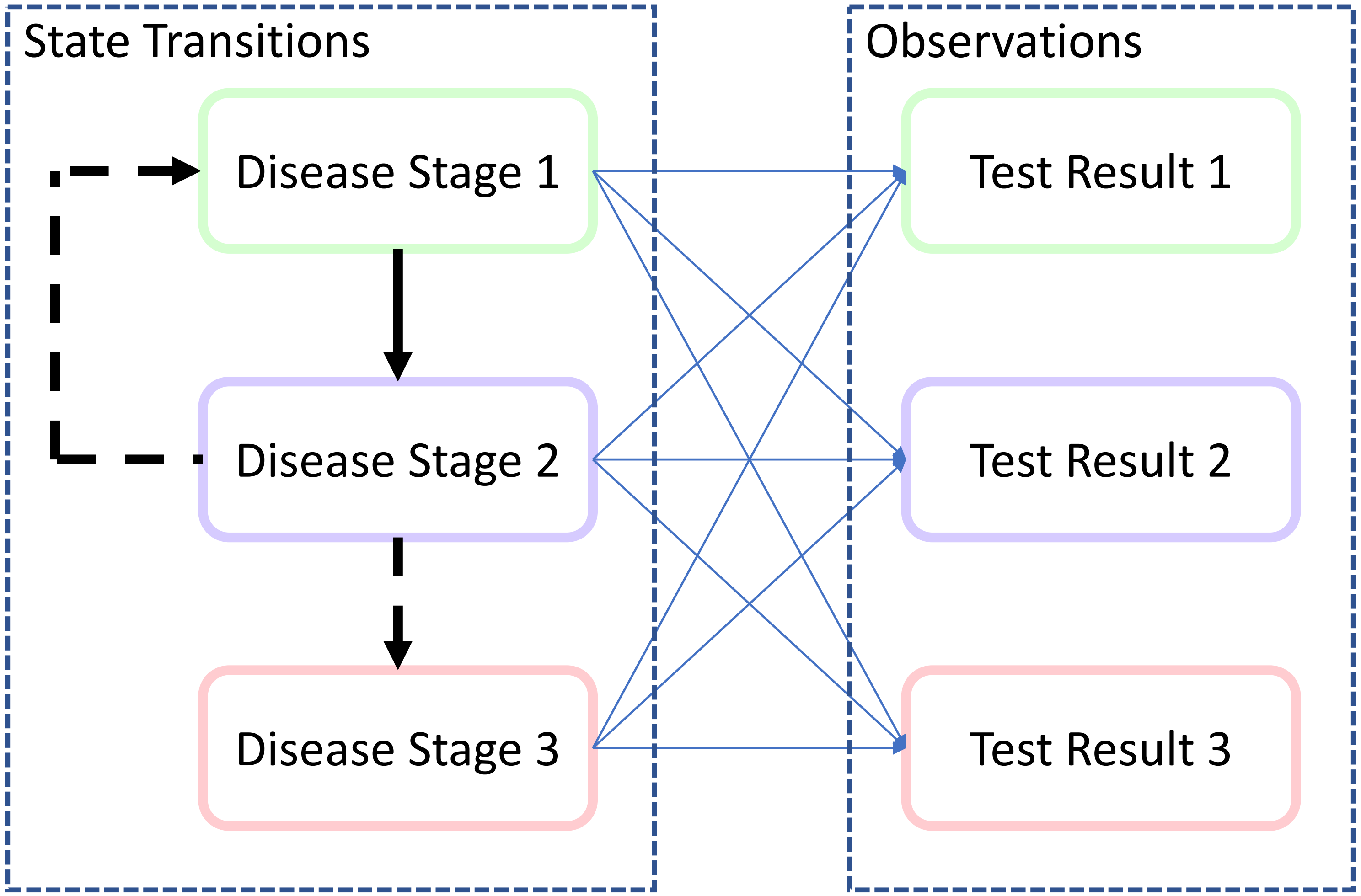
- a viral disease of the liver - preventable by vaccination
  - a number of stages of disease - treatment depends on stage
    - 240 million infected, 0.5-1.2 million die per year
- probabilistic progression between stages, depends on treatment
  - a controlled Markov model
    - Partially Observed Markov Decision Problem - POMDP
      - finite state, finite control action set, incomplete state observation
    - State observation depends on testing regime

the objective is health versus dollars

- costs for morbidity, treatment, testing
  - duality again - it costs money to test to save money

healthcare decision making via POMDP

simple POMDP based on hepatitis B disease progression and intervention



- three disease states - progressively more dire
- four decisions/controls: do nothing, inspect patient, order expensive but reliable test, apply costly treatment
- measurement outcomes: open-loop, unreliable test, reliable test
  - decisions can affect the information state - duality

POMDP stochastic optimal control

partially-observed Markov decision process

The partially-observed stochastic system becomes a Hidden Markov Model (HMM) with

$$\Pi_x(k+1) = \Pi_x(k) P(k, a_k), \quad \Pi(0)$$
$$\Xi_y(k) = \Pi_x(k) R(k, a_k)$$

- finitely enumerated state
- finitely enumerated action/control space
- finitely enumerated output space

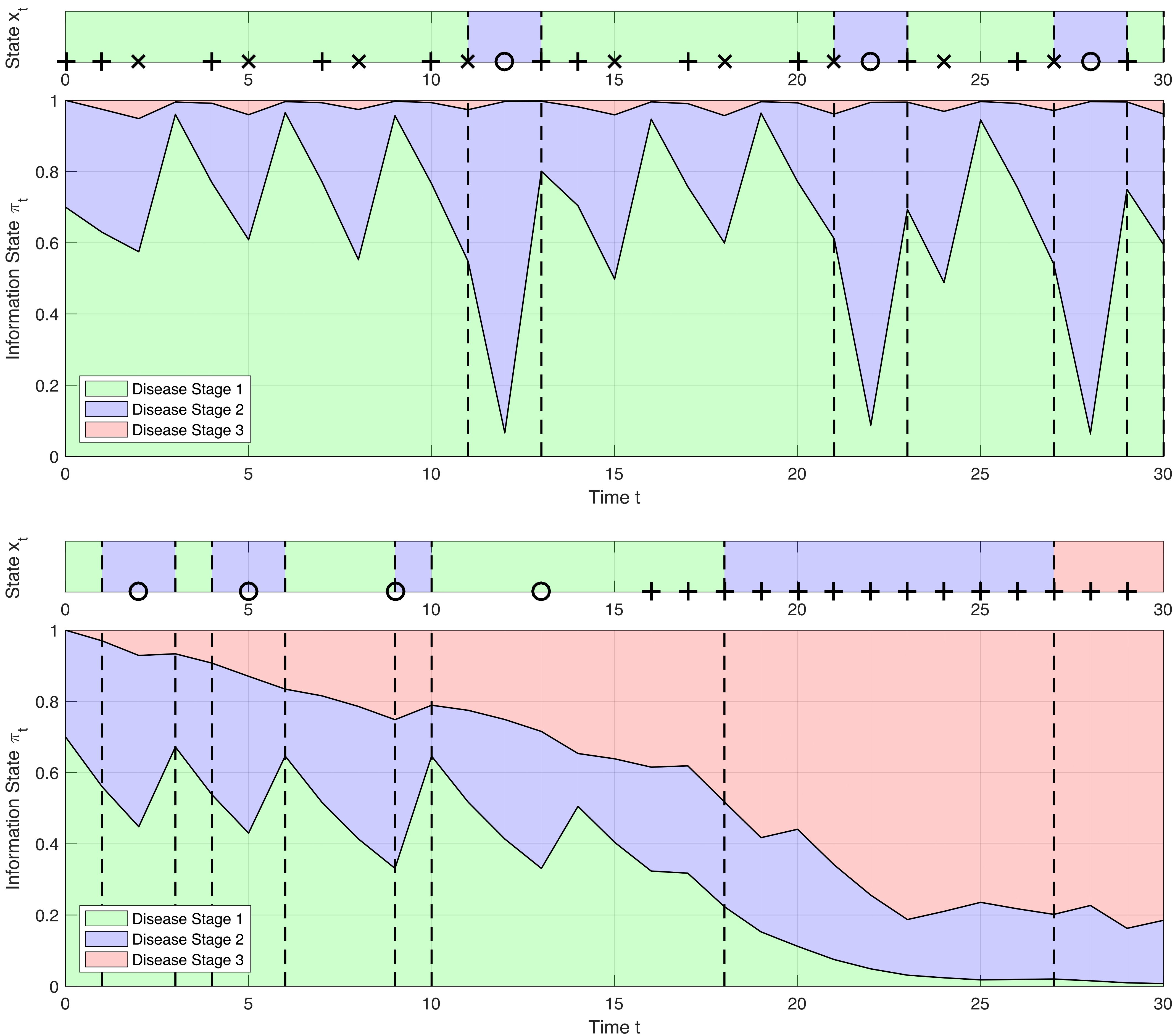
$$\mathbb{P}\left(x_{t+1} = j \mid x_t = i, u_t = a\right) = p_{ij}^a,$$
$$\mathbb{P}\left(y_{t+1} = \theta \mid x_{t+1} = j, u_t = a\right) = r_{j\theta}^a,$$

HMM state (Bayesian) filter
$$\pi_{t+1,j} = \frac{\sum_{i \in \mathbf{X}} \pi_{t,j} p_{ij}^a r_{j\theta}^a}{\sum_{i,j \in \mathbf{X}} \pi_{t,j} p_{ij}^a r_{j\theta}^a},$$

Finite-horizon stochastic output-feedback optimal control problem

$$V_k(\pi_k) = \min_{u \in \mathbf{U}} \left\{ \pi_k c(u) + \alpha \sum_{y \in \mathbf{Y}} \mathbb{P}\left(y \mid \pi_k,\ u\right) V_{k+1}(\pi_{k+1}) \right\}, V_N(\pi_N) = \pi_N c_N.$$

dual optimal vs certainty equivalence control



N = 6
+ appmnt
o treatment
x test

dual stochastic MPC

certainty equivalence stochastic MPC

Message 4

medical testing as part of treatment is dual control
it helps to focus on the information state

it costs money now
to save money in the future

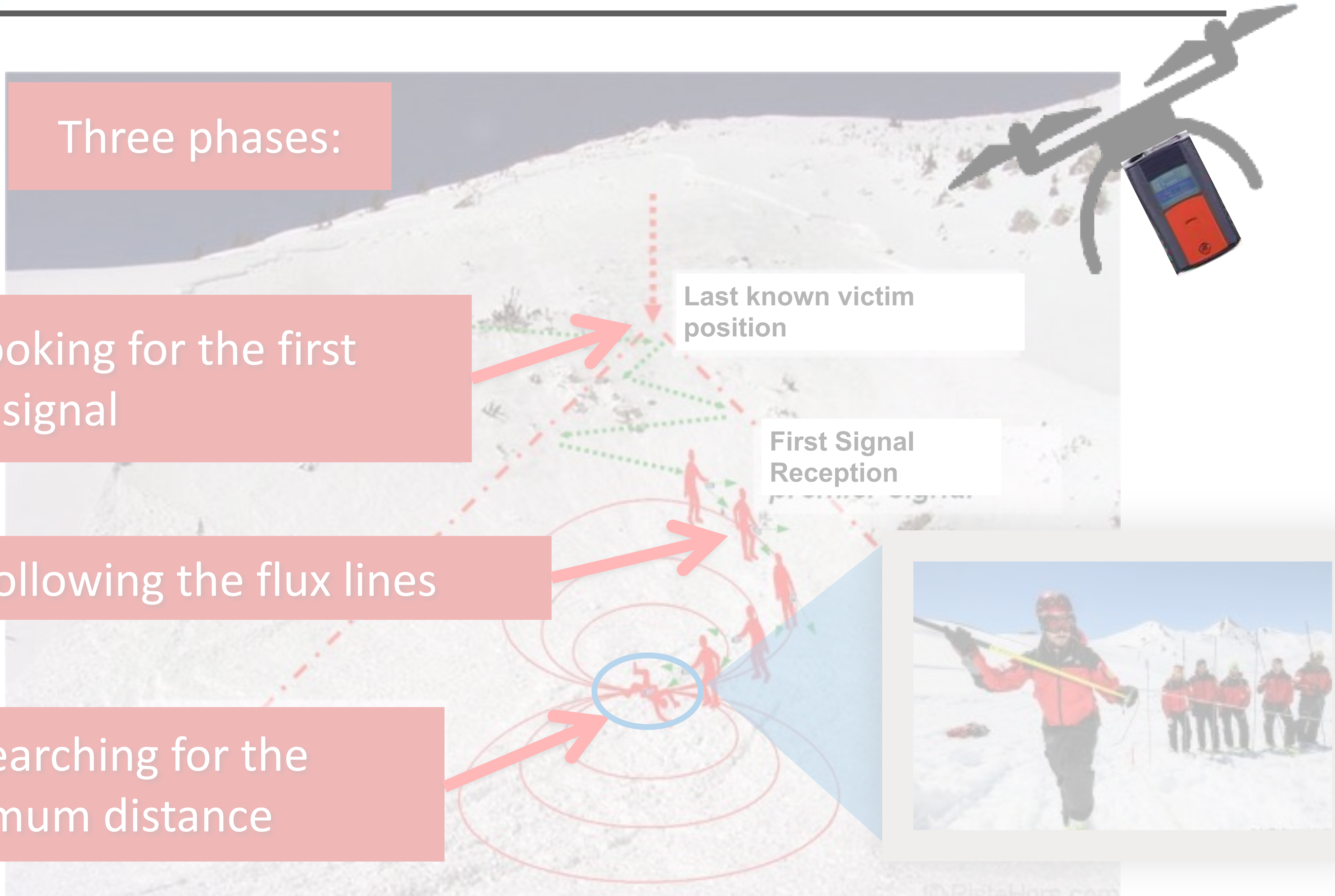
Part 5: autonomous systems and learning

how hard could it be?

Four slides used with kind permission from Lorenzo Marconi's
American Control Conference plenary July 12, 2019
"Aerial Robotics: challenges and opportunities outside the lab

A.R.V.A.
Appareil de Recherche de Victims d'Avalanche

Three phases:

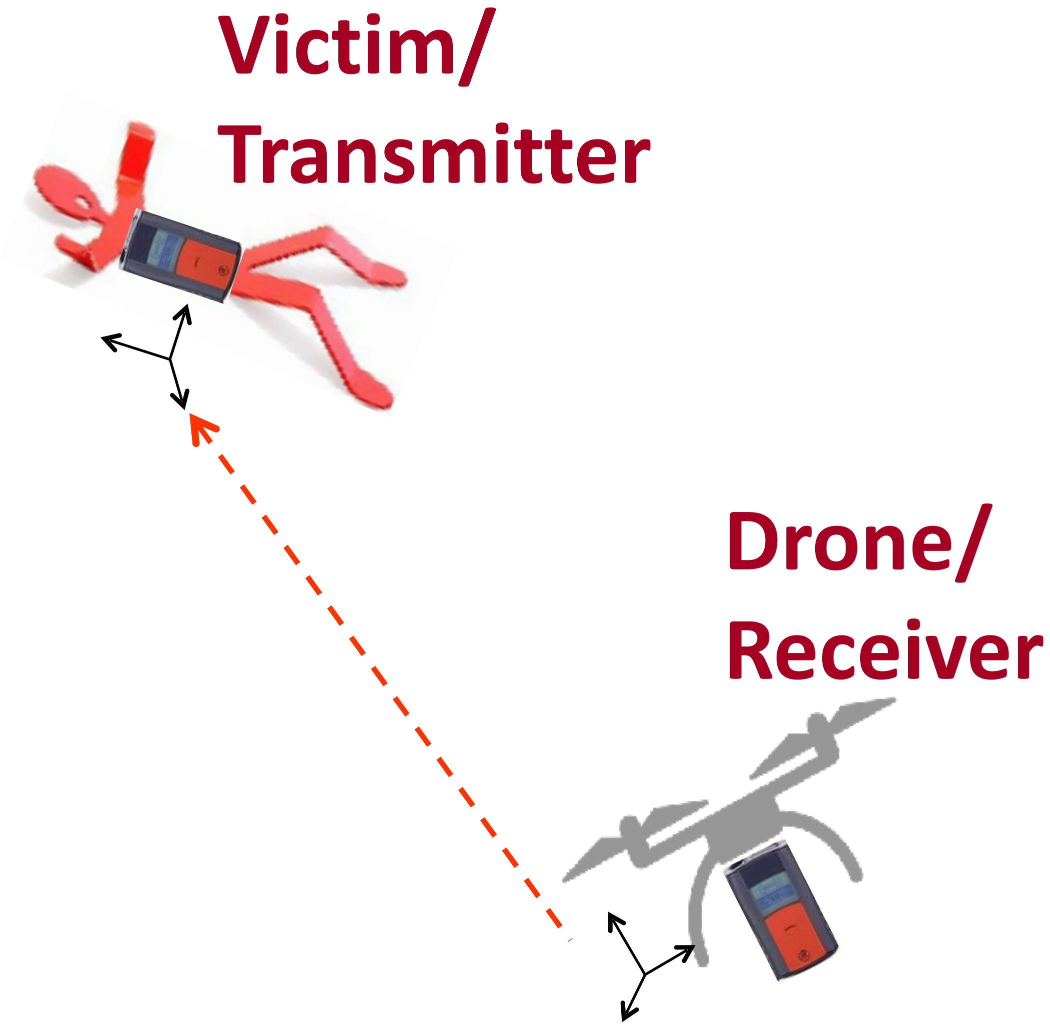


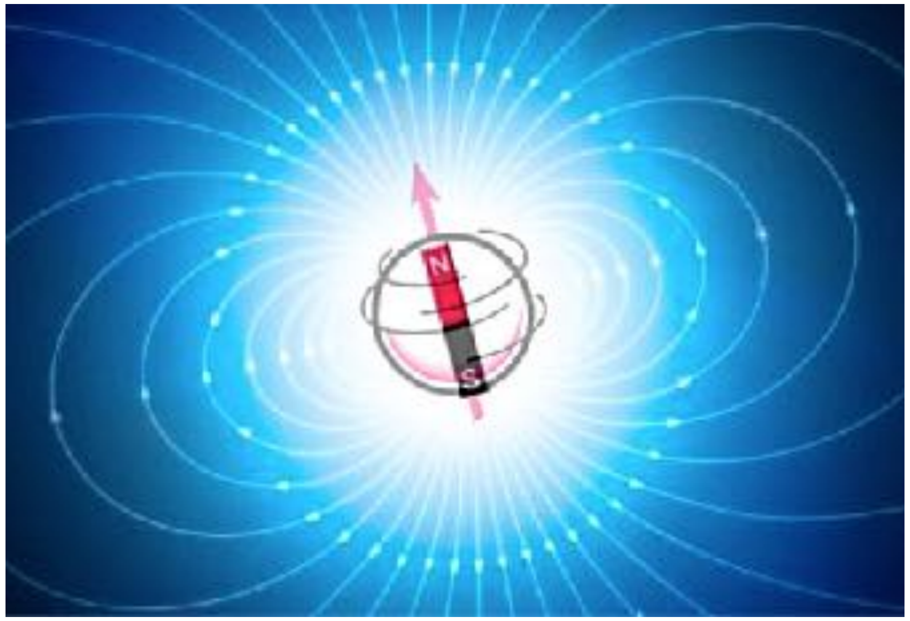
1st Looking for the first valid signal

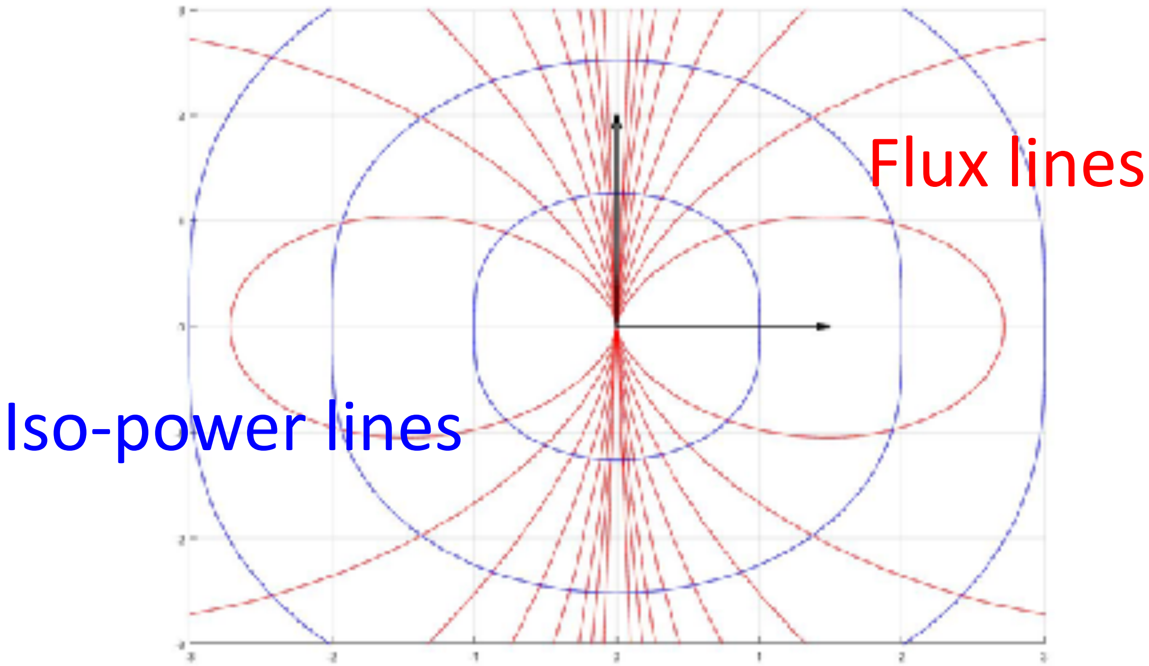
2nd Following the flux lines

3rd Searching for the minimum distance

The magnetic field





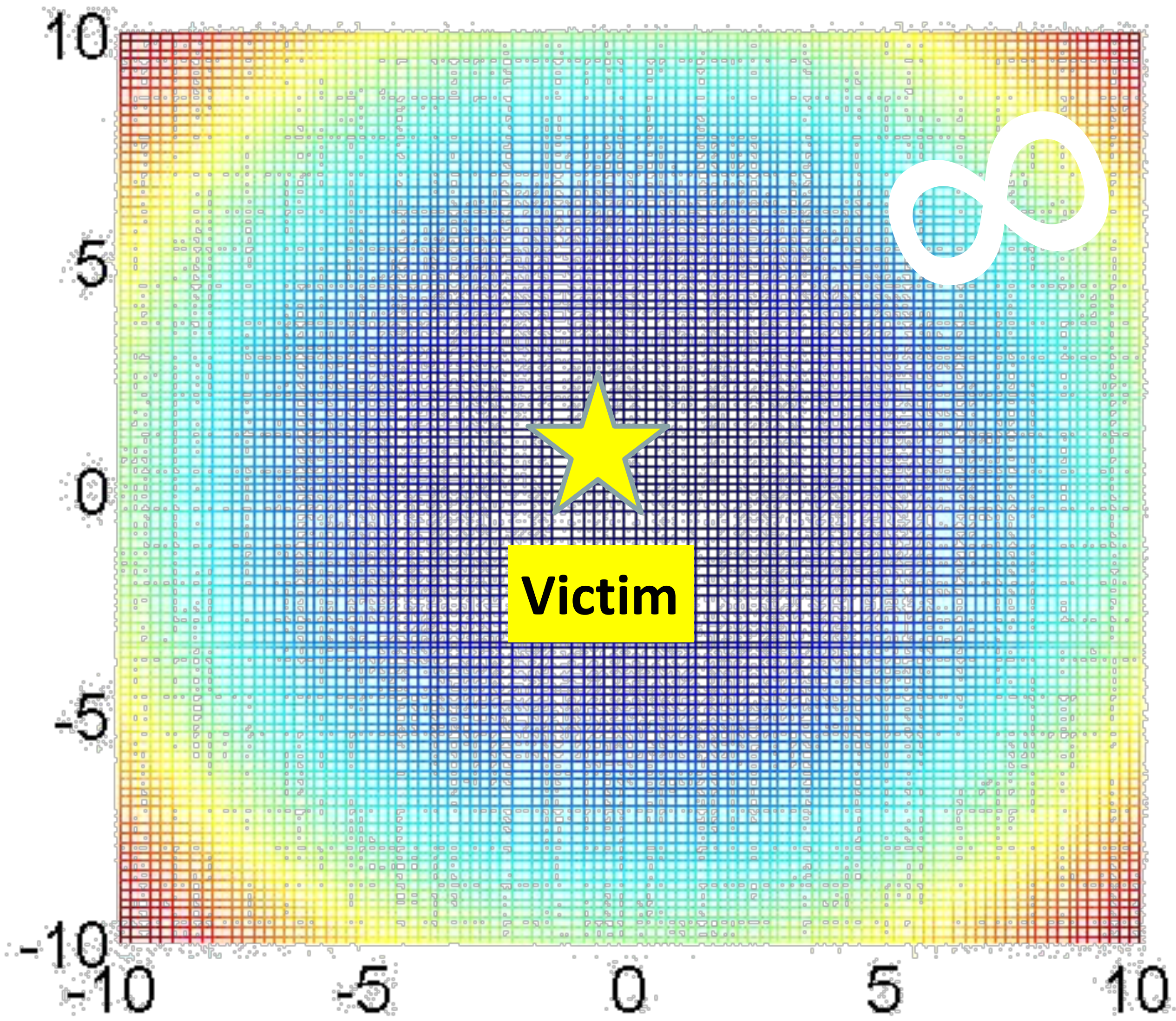


[modified from original slide]

Projection of the flux lines strongly dependent on the victim orientation

A.R.V.A.: Observer

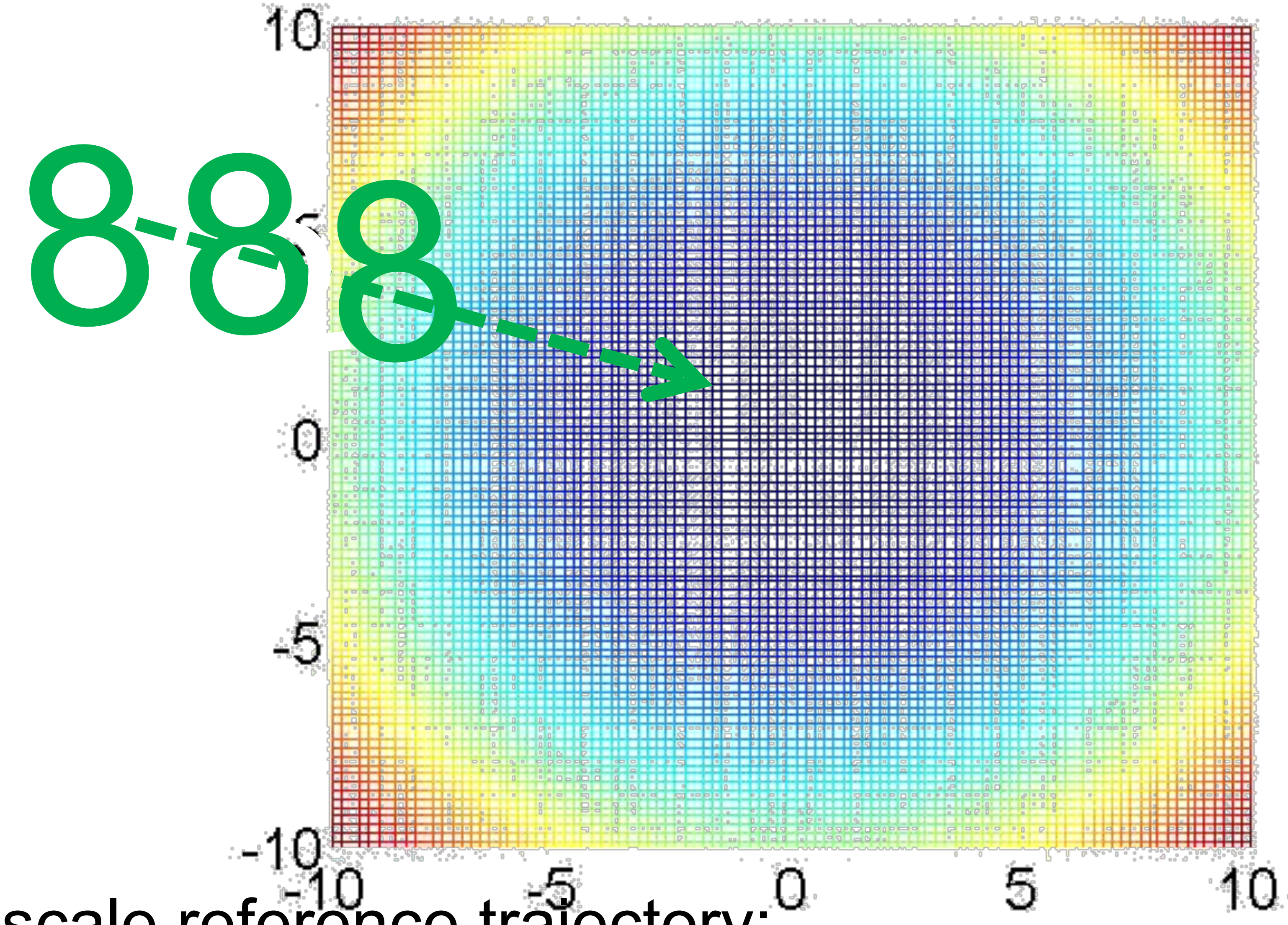
- Loitering maneuvers to excite the system (observability)
- Observer design
  - Kalman filter,
  - Recursive Least Squares
  - Luenberger-like observers



N. Mimmo. P. Bernard, L. Marconi, in preparation

A.R.V.A.: Gradient descend methods

Extremum seeking algorithm



Two time-scale reference trajectory:

- “High-frequency” loitering for estimation of the distance gradient
- “Low-frequency” trajectory straight to the victim

Autonomous systems

Elevators as autonomous systems

1924 Elevators without attendants

More challenging environments require more sophisticated sensing

exploration vs exploitation tradeoff

well understood in robotics

not so well in vehicles … yet

Even suboptimal control in uncertain environments requires devotion of control resources to sensing the environment

This comes at a cost to control compared with operating in a known environment

Without it the systems fail

Message 5

There seem to be two distinct aspects of autonomy

Either

- operate in a highly regulated environment, or
- devote a significant effort to situational awareness
  - deliberate maneuvers to improve observability
  - multiple and mobile sensors

Or maybe operate in distinct modes

Conclusions

Agile stochastic control systems without full state information require the control signal to provide persistent probing to facilitate state estimation

This comes at a cost to the estimation-free control performance

The control signal balances regulation and excitation - *duality*

It is inherently part of stochastic optimal control and is very hard to optimize computationally

Three examples of commercial suboptimal stochastic control ~~will be~~ were presented - mobile wireless, TCP, hepatitis B management

Each contains dual action

expending energy to save energy

consuming capacity to save capacity

spending money to save money

Settle for suboptimal control

but do not forget the excitation requirements

References - a biased sample

Automatica 47 (2011) 65–78

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Stochastic observability in network state estimation and control☆

Andrew R. Liu*, Robert R. Bitmead

Department of Mechanical & Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA

Automatica 74 (2016) 84–89

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Optimal dual adaptive agile mobile wireless power control☆

Minh Hong Ha, Robert R. Bitmead

Department of Mechanical & Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla CA, 92093-0411, USA

Automatica 94 (2018) 315–323

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Stochastic output-feedback model predictive control☆

Martin A. Sehr [a], Robert R. Bitmead [b],*

[a] Siemens Corporate Technology, Berkeley, CA 94704, USA

[b] Department of Mechanical & Aerospace Engineering, University of California, San Diego, La Jolla, CA 92093-0411, USA

Handbook of Model Predictive Control

Saša V. Raković • William S. Levine

Editors

Birkhäuser

Probing and Duality in Stochastic Model Predictive Control

Martin A. Sehr and Robert R. Bitmead

Sehr *et al. Theoretical Biology and Medical Modelling* (2017) 14:11

DOI 10.1186/s12976-017-0057-6

Theoretical Biology and Medical Modelling

RESEARCH Open Access

Markov modeling in hepatitis B screening and linkage to care

Martin A. Sehr[1], Kartik D. Joshi[2], John M. Fontanesi[3], Robert J. Wong[4], Robert R. Bitmead[1] and Robert G. Gish[5,6,7*]

2017 IEEE Conference on Control Technology and Applications (CCTA)

August 27-30, 2017. Kohala Coast, Hawai'i, USA

Tractable Dual Optimal Stochastic Model Predictive Control: An Example in Healthcare

Martin A. Sehr & Robert R. Bitmead

Thank you