

Probing the Temporal Characteristics of Available Bandwidth

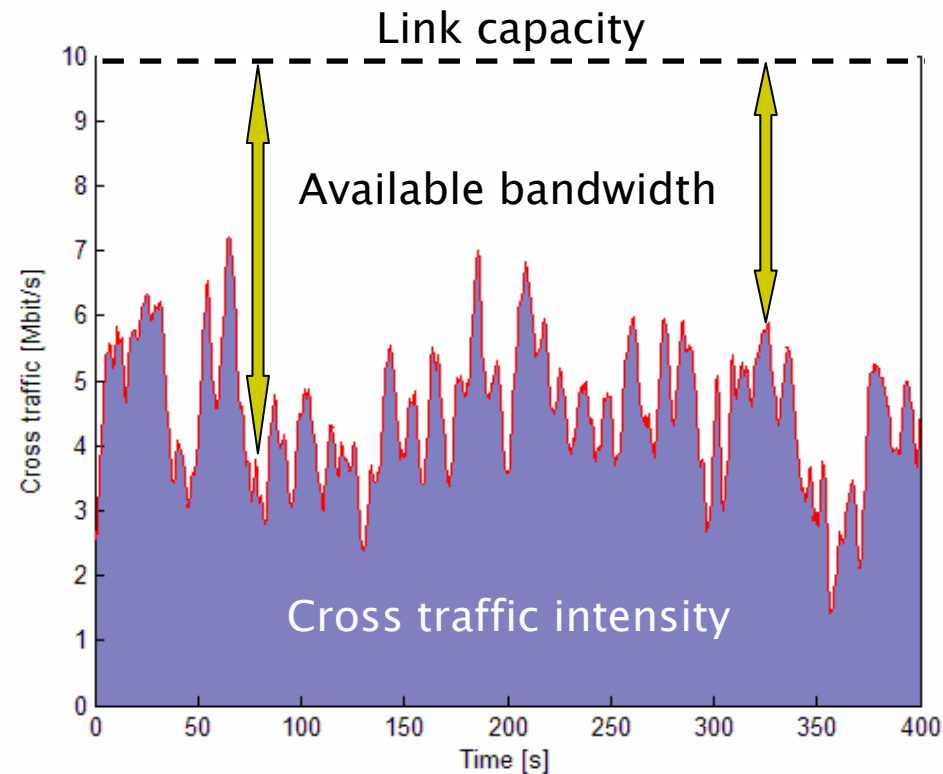
Erik Hartikainen
Industrial Ph.D. Student

Outline

- Measuring available bandwidth
- BART – Bandwidth Available in Real-Time
- Kalman Filtering
- The Process Noise Covariance (Q)
- Experiments – Probing the Temporal Characteristics
- Summary

Measuring available bandwidth

- What is available bandwidth?
 - The available bandwidth on a network link is a measure of the unutilized capacity



Measuring available bandwidth

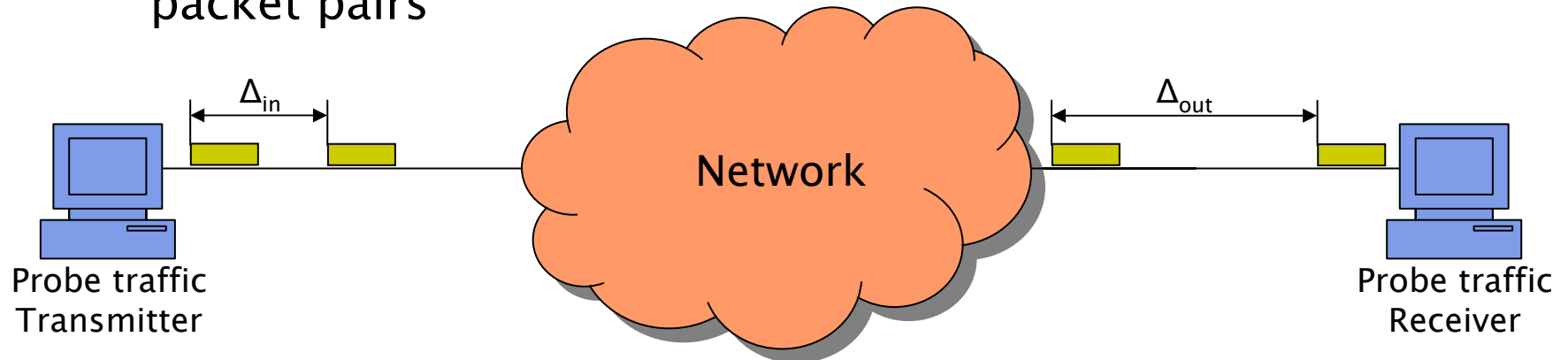
- Why measuring available bandwidth?
 - Network monitoring
 - Server selection
 - Verifying service level agreements
 - Tuning of audio/video codec
 - ...

BART – Bandwidth Available in Real-Time

- Characteristics of BART
 - Active probing
 - Kalman filtering
 - Real-time estimation
 - Tuning possibilities:
 - The probe packet size
 - The number of probe-packet pairs in each measurement
 - The organization of the probe pairs (separated or as a train)
 - The inter-measurement time separation
 - The probability distribution for the probing intensity
 - The covariance matrix of the process noise (Q)

BART – Bandwidth Available in Real-Time

- Active probing
 - Alternate the probe traffic intensity u
 - Use timestamps to compute the strain between probe packet pairs

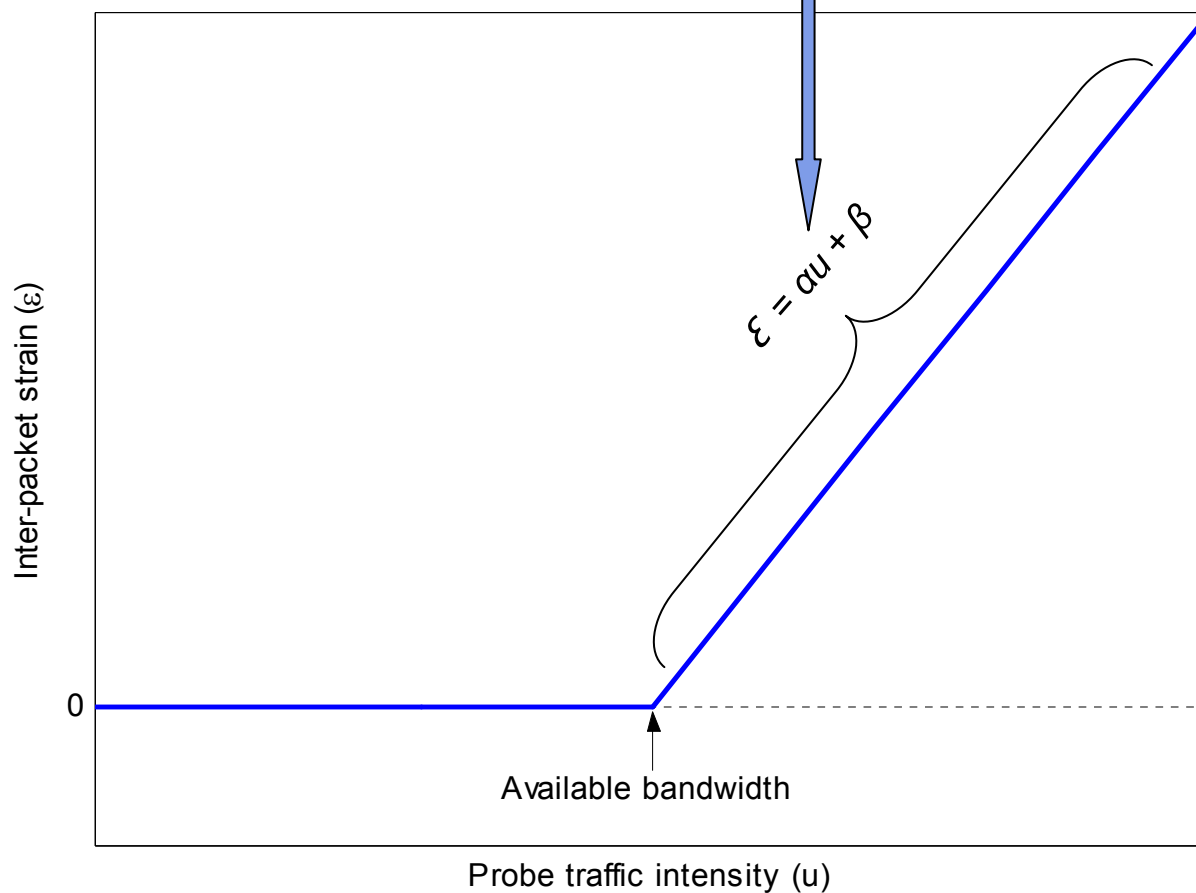


- $\mathcal{E}_i =$ relative time dispersion of probe pair $i = \frac{\Delta_{out}^i - \Delta_{in}^i}{\Delta_{in}^i}$
- The average strain \mathcal{E} of N probe packet pairs serves as an input parameter to the Kalman filter

BART – Bandwidth Available in Real-Time

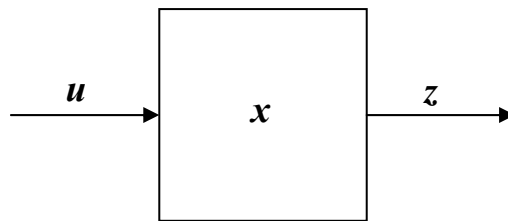
- Active probing

[Use Kalman filtering to estimate α and β !]



Kalman Filtering

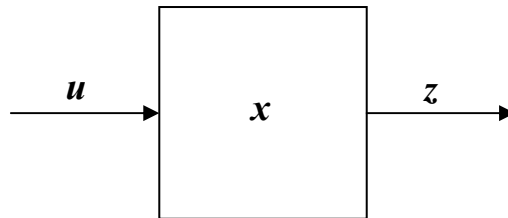
- We have a system which is described by a state vector x
- We can affect the system with an input u
- We can only observe the system through measurements z



- The evolution of the system state: $x_k = f(x_{k-1}) + w_{k-1}$
- The measurements: $z_k = h(x_k) + v_k$
 - The functions f and h represent the system evolution model and the measurement model, respectively
 - w is the process noise and v is the measurement noise

Kalman Filtering

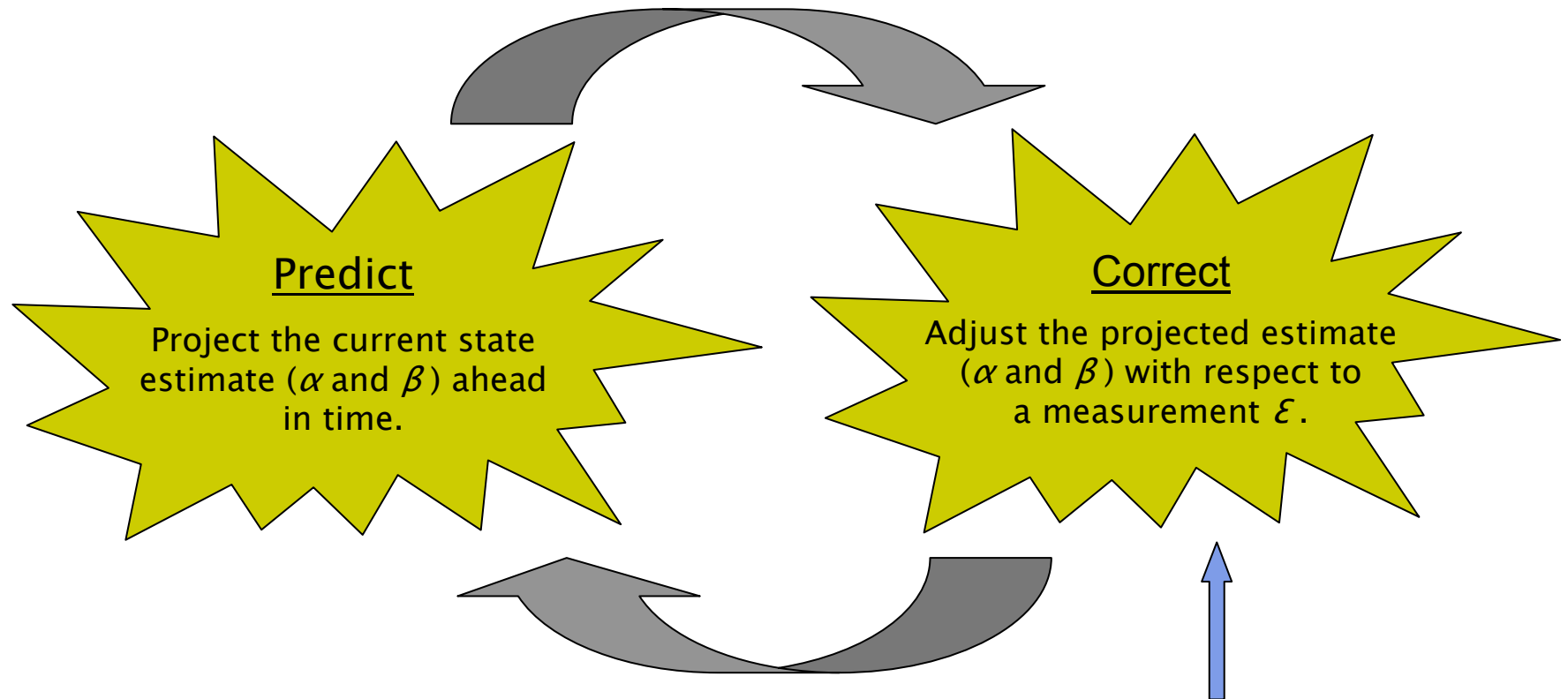
- The Kalman filter is known to be the optimal filter if:
 - f and h are linear functions
 - The noise w and v are white Gaussian
- Kalman filters often work well, even for non-Gaussian noise and slightly non-linear models



- In BART:
 - The state vector x consists of α and β
 - The input u is the probe traffic
 - The measurement z is the observed inter-packet strain \mathcal{E}

Kalman Filtering

- Kalman filtering – an overview

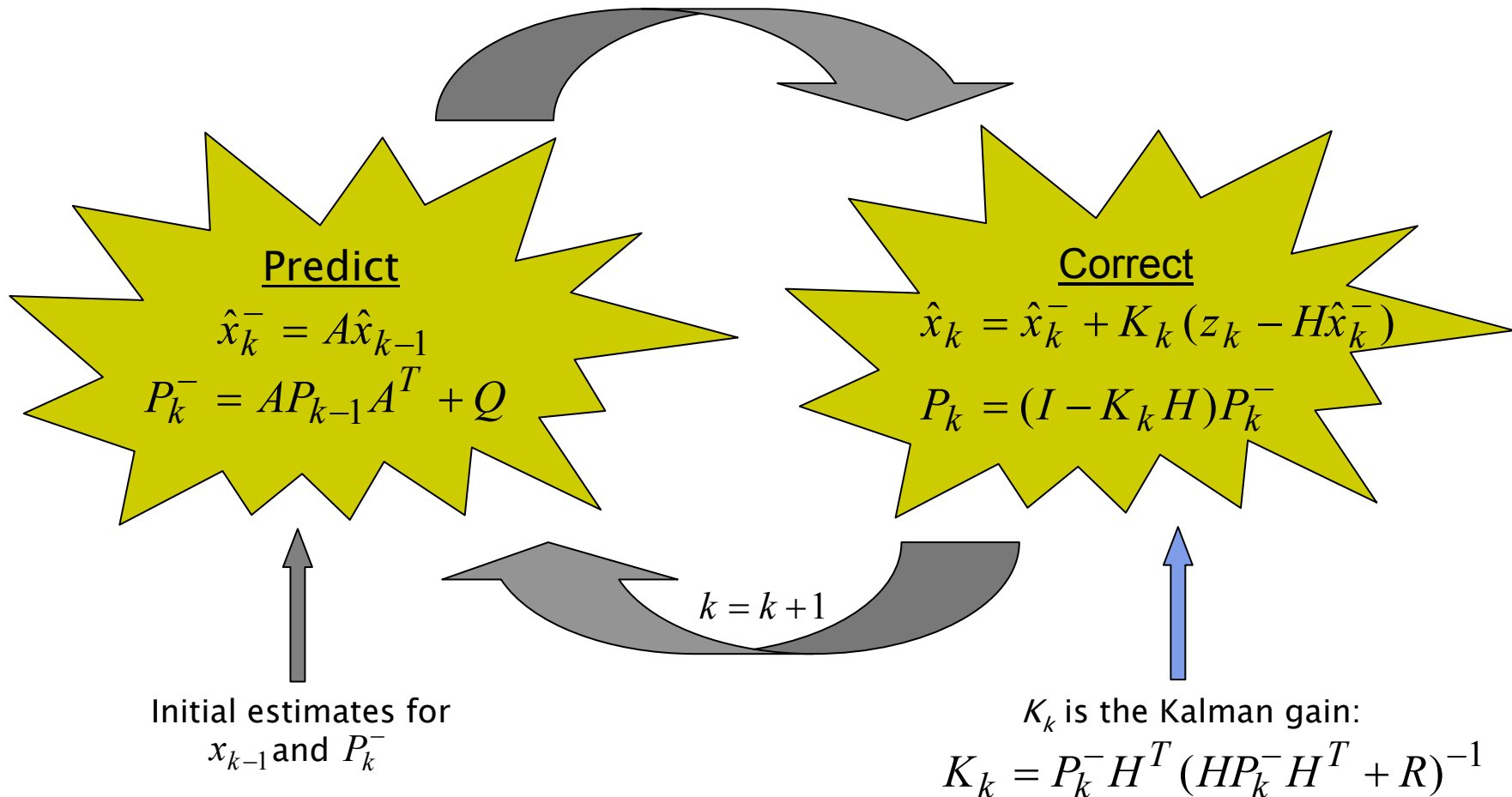


Of great importance:

The precision of the measurement (R)
The expected variations in the system (Q)

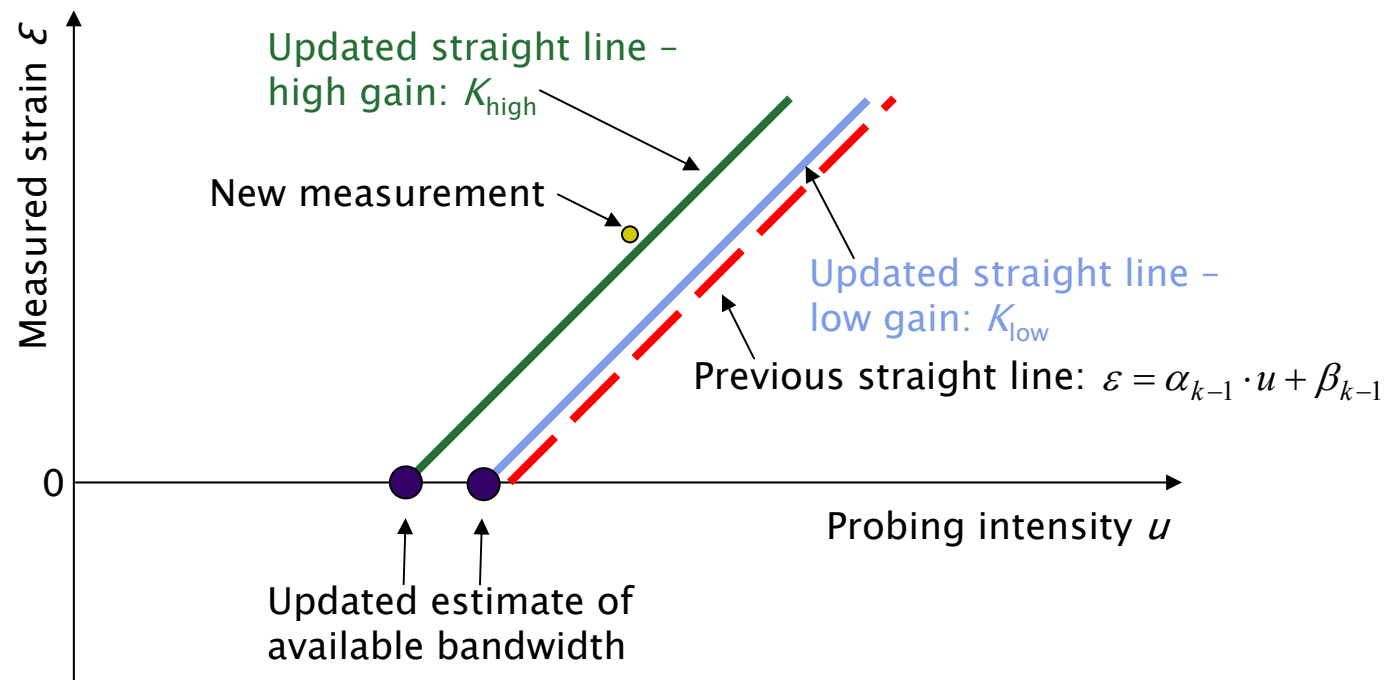
Kalman Filtering

- Kalman filtering – the equations



The Process Noise Covariance (Q)

- The Kalman gain K is important since it describes the relative weight given to the new measurement



- The values of Q are important since Q affects the Kalman gain

The Process Noise Covariance (Q)

- In BART, the evolution of the system state can be written as:

$$x_k = x_{k-1} + w_{k-1} \quad \Leftrightarrow \quad \begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} = \begin{bmatrix} \alpha_{k-1} \\ \beta_{k-1} \end{bmatrix} + w_{k-1}$$

- The process noise w is defined as the change of the system state between two consecutive measurements:

$$w = \Delta x = \begin{bmatrix} \Delta\alpha \\ \Delta\beta \end{bmatrix}$$

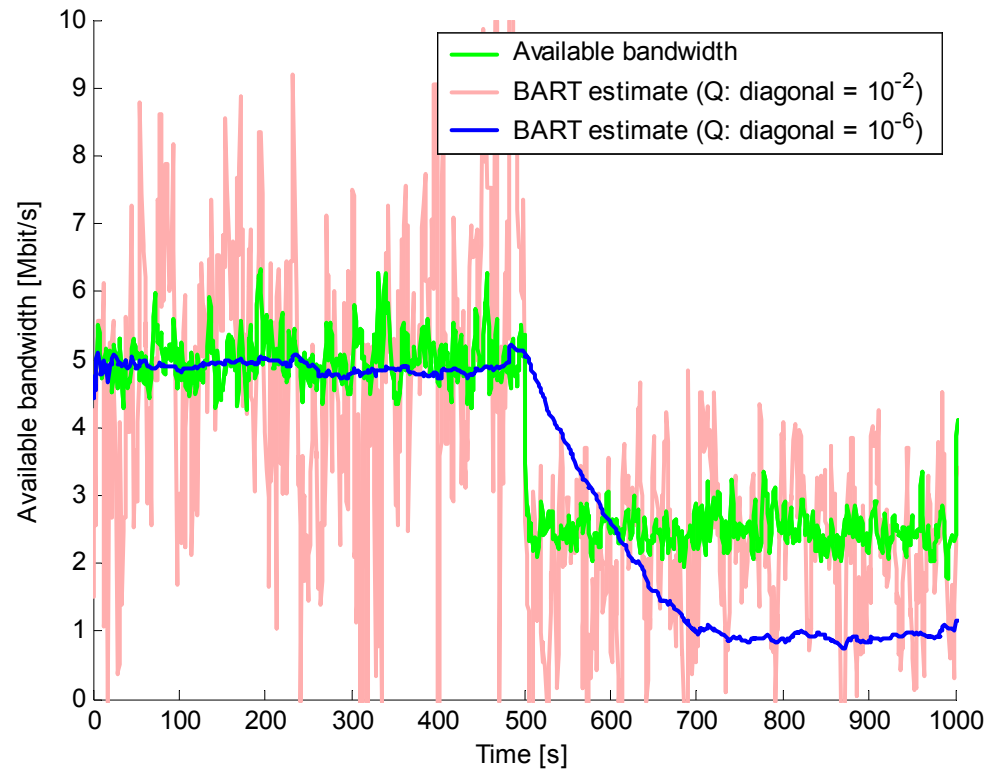
- So, the process noise covariance becomes:

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} V(\Delta\alpha) & C(\Delta\alpha, \Delta\beta) \\ C(\Delta\alpha, \Delta\beta) & V(\Delta\beta) \end{bmatrix}$$

$$(Q_{11} \geq 0, \quad Q_{22} \geq 0, \quad -\sqrt{Q_{11} Q_{22}} \leq Q_{12} \leq \sqrt{Q_{11} Q_{22}}, \quad Q_{21} = Q_{12})$$

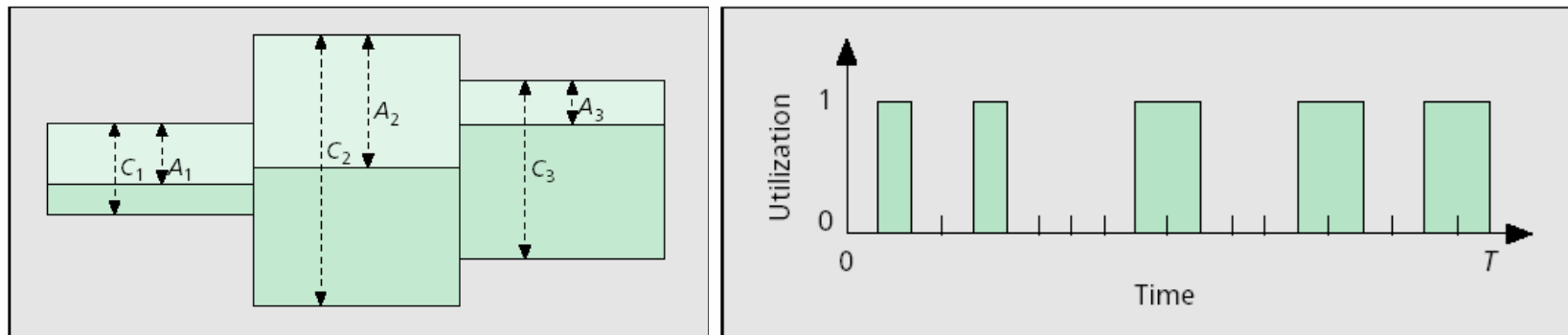
The Process Noise Covariance (Q)

- The values of Q affects the characteristics of BART
- Stability can be traded for agility
- Two extreme cases:



The Process Noise Covariance (Q)

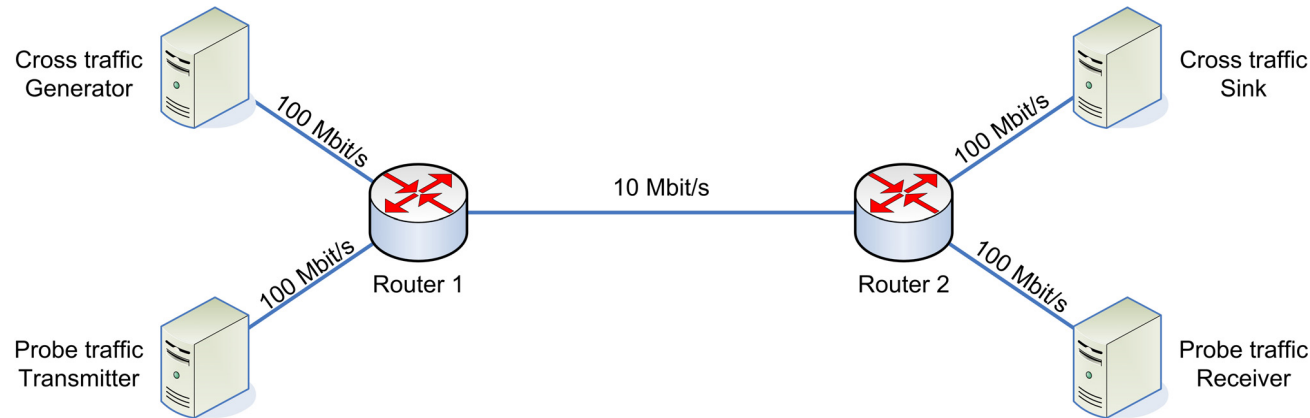
- Stability vs. agility:
 - The choice depends on the characteristics of the true available bandwidth
- The available bandwidth is dependent on both the link capacity and the link load



- The traffic load (and thus also the available bandwidth) is often defined relative some time resolution τ
- Two factors affecting the available bandwidth characteristics:
 - The time resolution and the cross-traffic aggregation

Experiments

- Optimization of Q with respect to different time resolution and cross-traffic aggregation



- True available bandwidth:
 - Averaging time scale $\tau = 1, 2, 4, 8, 16, 32, 64, 128$ seconds
- Traffic cases:
 - Synthetic cross traffic: approximately 10 and 100 active users

Experiments

- What should we expect?

- A single bottleneck link ->

No/small variations in α ->

$Q(1,1)$ should be small - recall: $Q_{11} = V(\Delta\alpha)$, $\alpha = 1/C_{\text{link}}$

- Variations in cross-traffic intensity ->

Variations in β proportional to variations in cross-traffic intensity ->

$Q(2,2)$ should be different depending on time resolution and/or cross-traffic aggregation level - recall: $Q_{22} = V(\Delta\beta)$, $B = -\beta/\alpha$

Small time resolution and/or low aggregation level ->

Large variations in β -> $Q(2,2)$ should be large

Large time resolution and/or high aggregation level ->

Small variations in β -> $Q(2,2)$ should be small

- No obvious and connection between variations in α and variations in β ->

$Q(1,2) = Q(2,1)$ should be small - recall: $Q_{12} = Q_{21} = C(\Delta\alpha, \Delta\beta)$

Experiments

- In order to tune the temporal characteristics of the estimation:

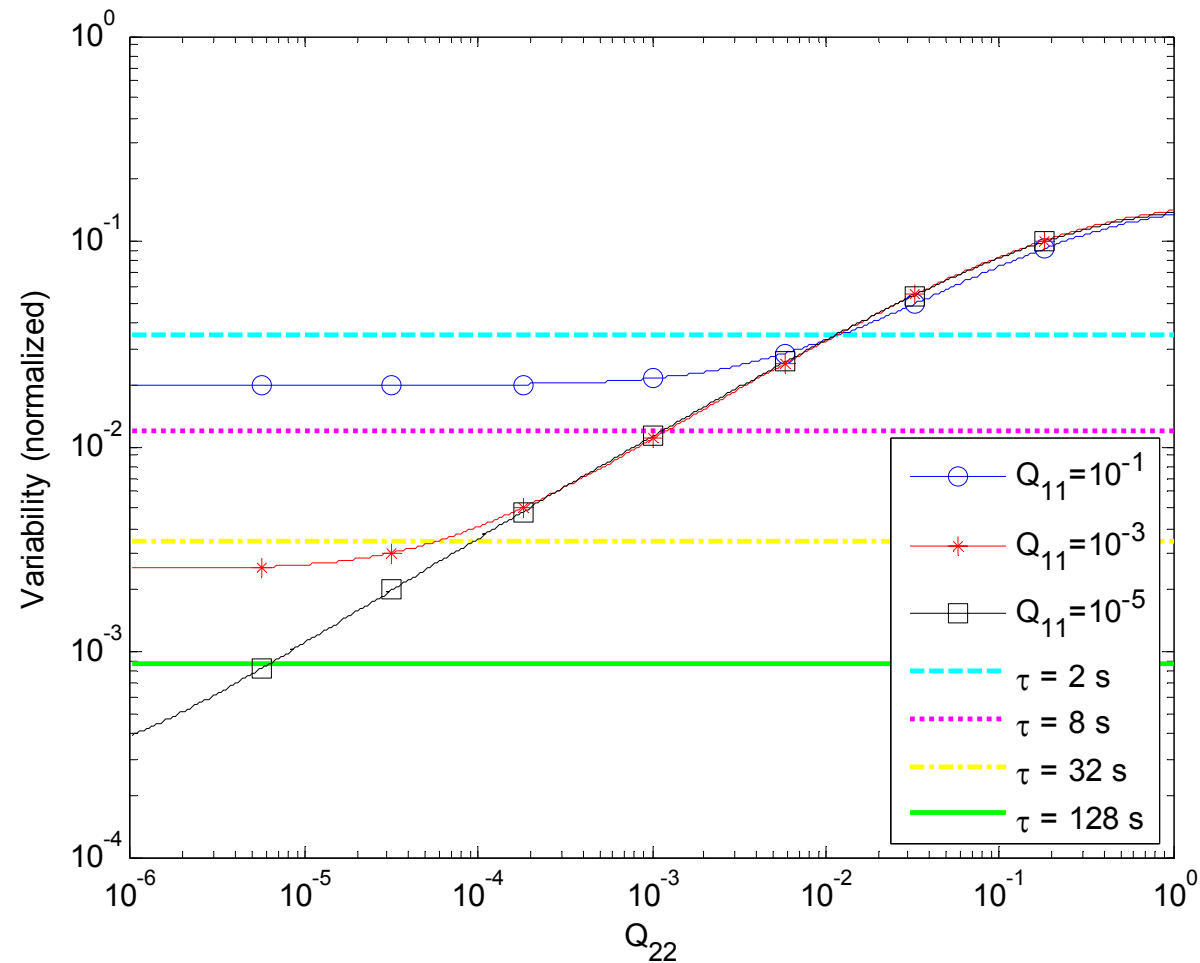
$$\text{Variability} = \sqrt{V\{\Delta B_k\}} = \sqrt{V\{B_2 - B_1, B_3 - B_2, \dots\}}$$

$$\left(B = -\beta/\alpha, \quad Q_{22} = V(\Delta\beta) \right)$$

- We expect an improved performance of BART if the variability of the estimates coincide with the cross-traffic variability

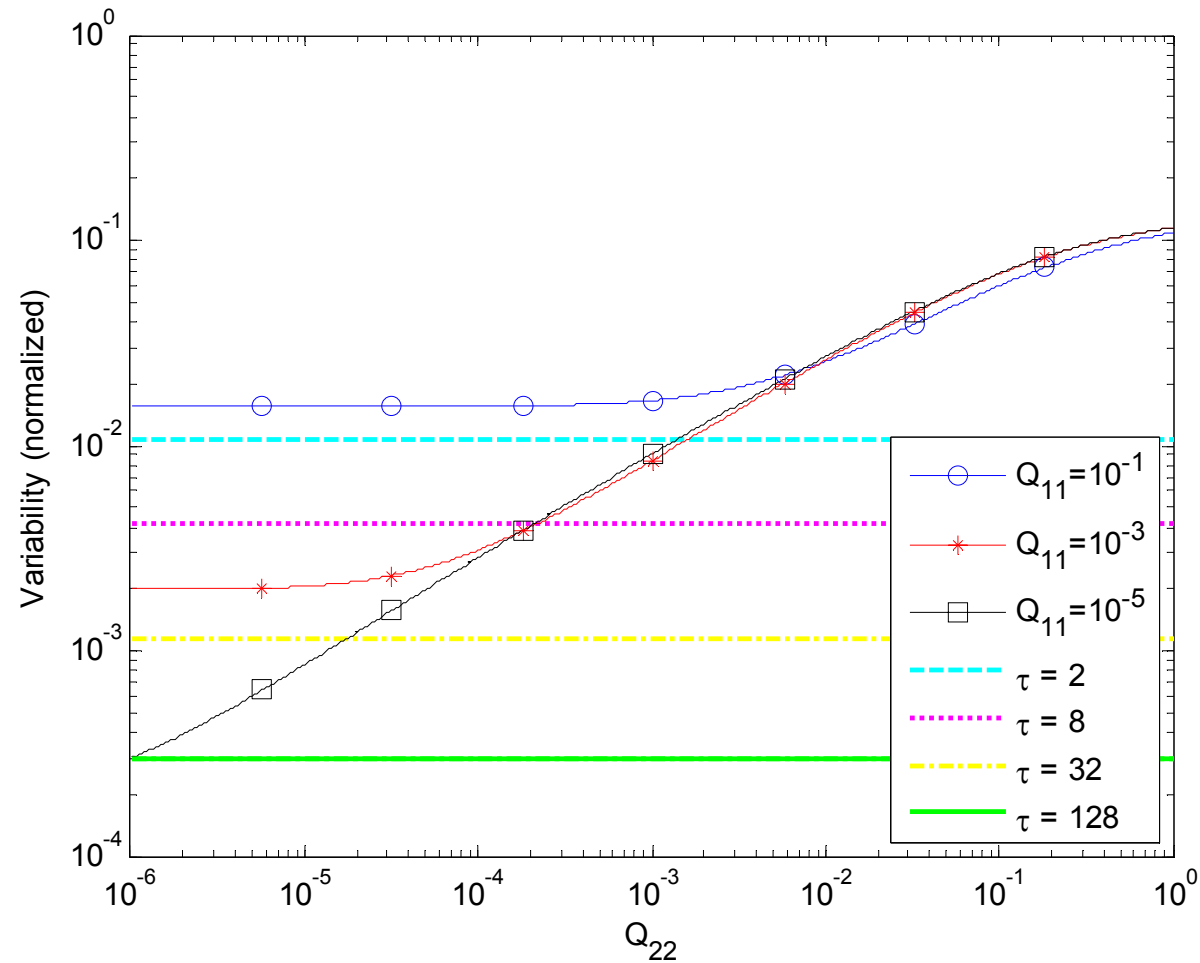
Experiments

- Approx. 10 active users (low aggregation):



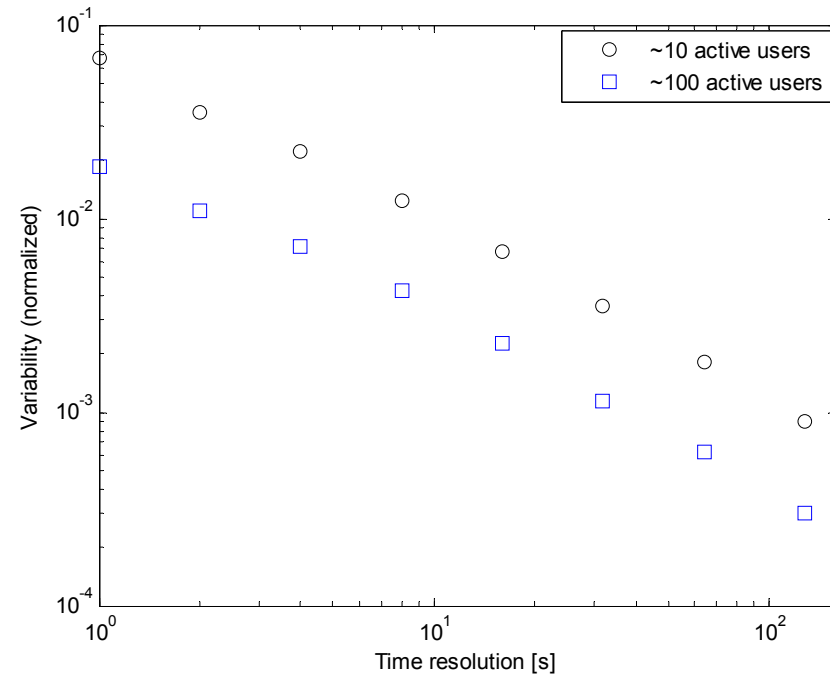
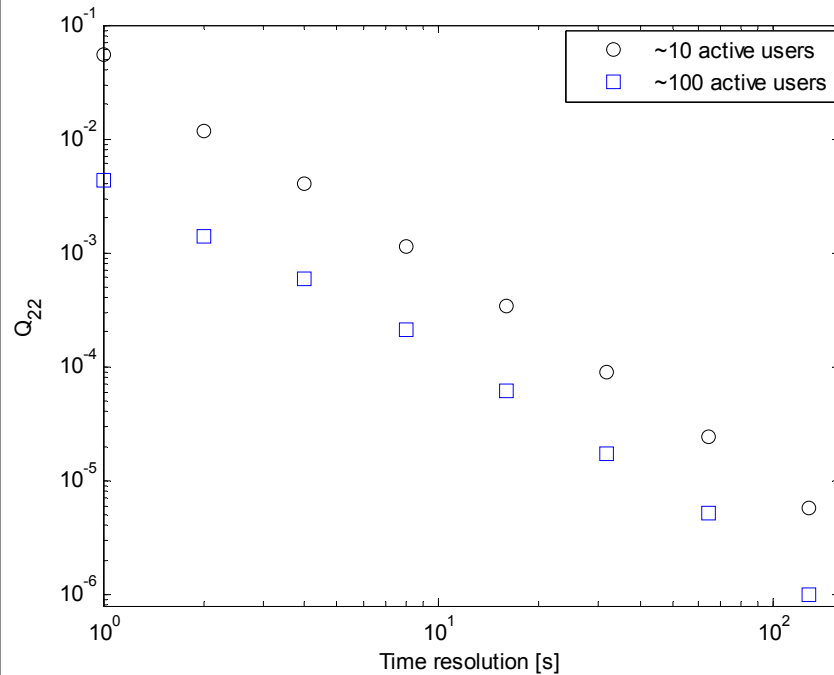
Experiments

- Approx. 100 active users (high aggregation):



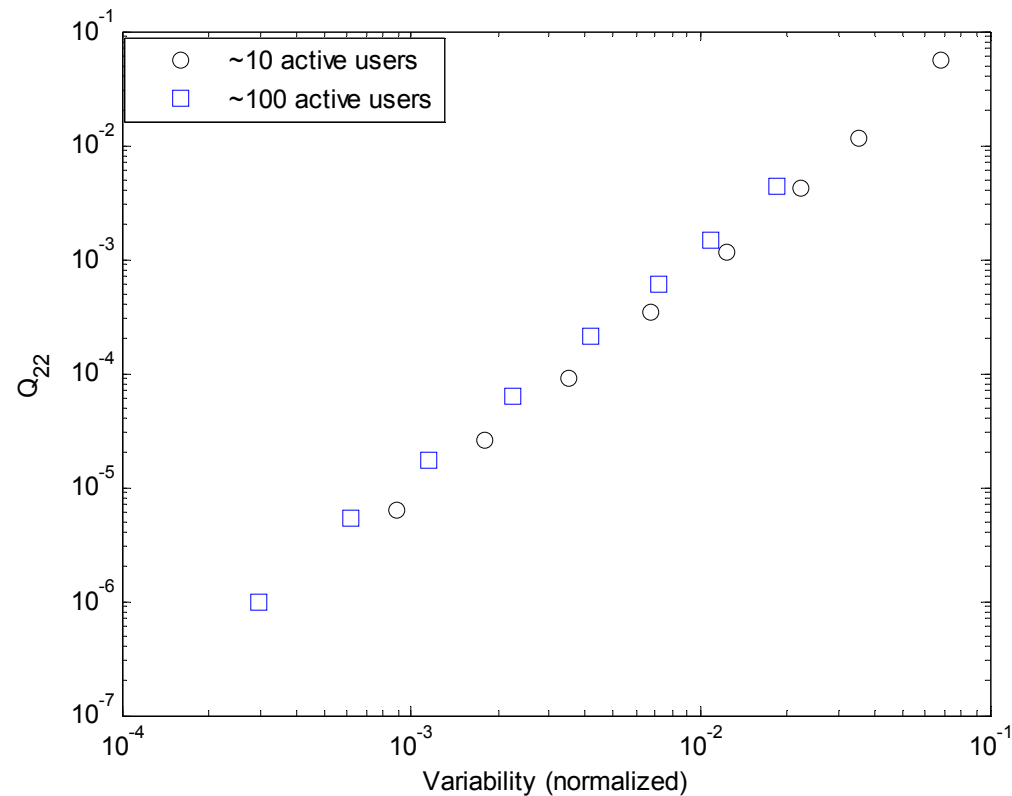
Experiments

- A power-law behavior is observed:
 - Optimal $Q(2,2)$ vs. time resolution (left figure)
 - Cross-traffic variability vs. time resolution (right figure)



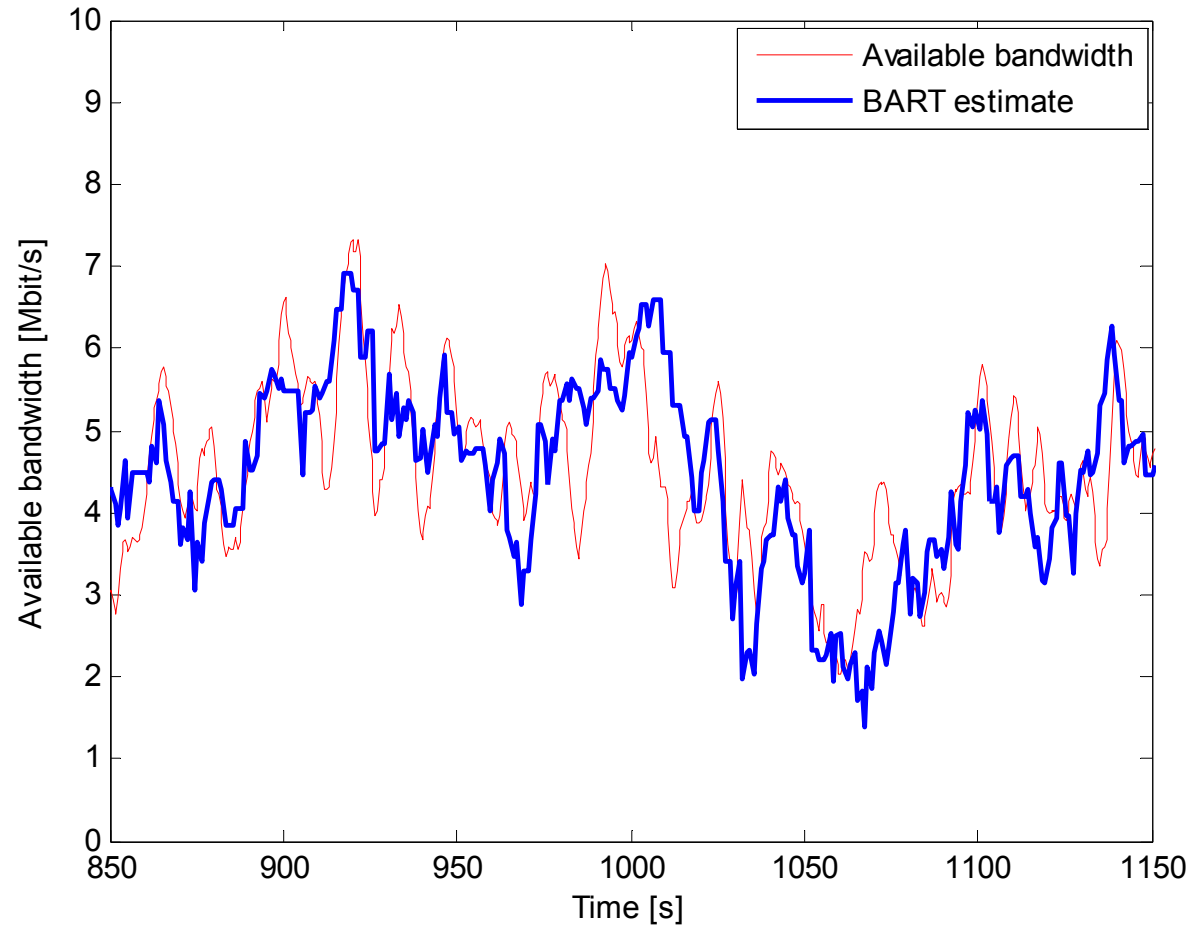
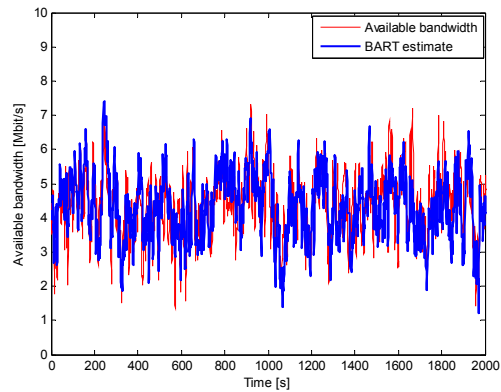
Experiments

- The correlation between the optimal $Q(2,2)$ and the variability of the cross traffic:



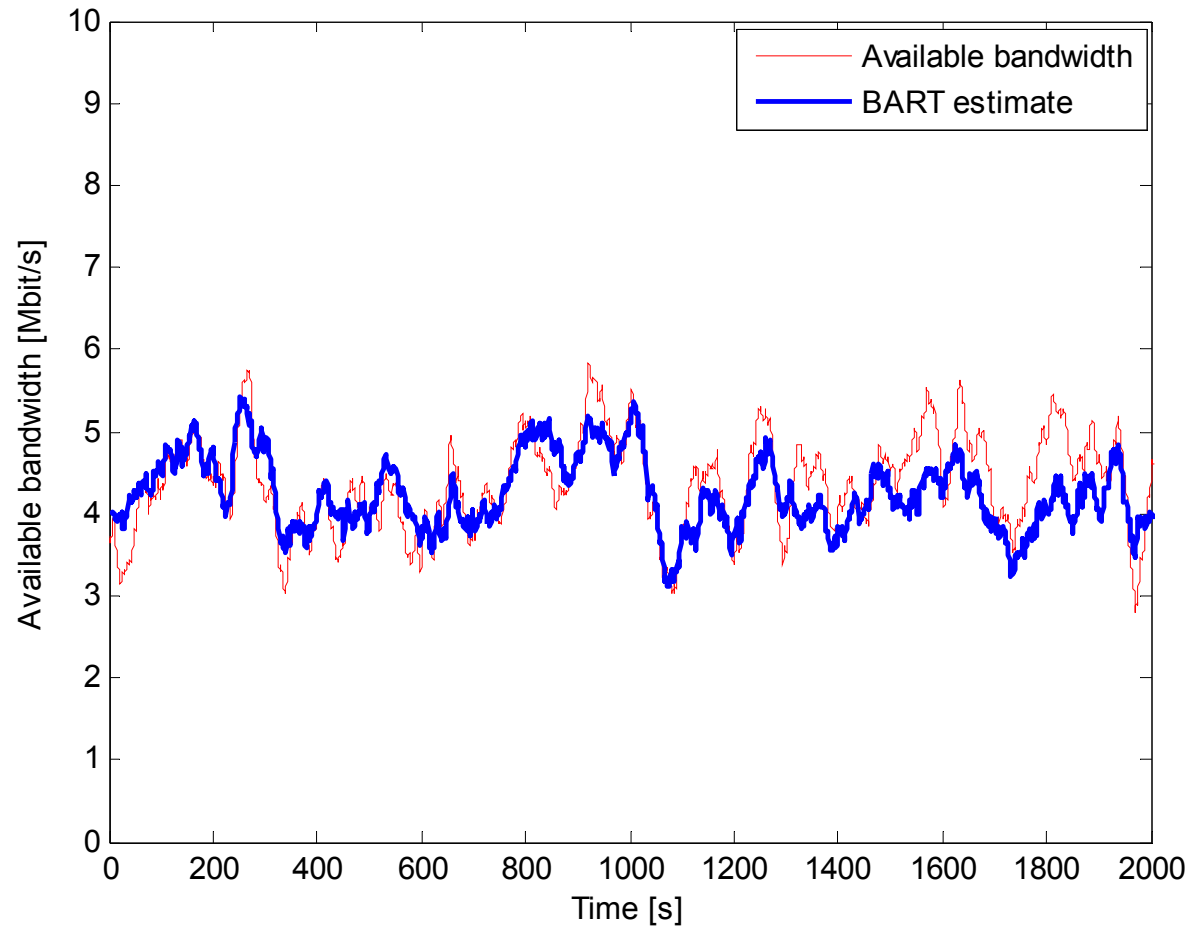
Experiments

- BART estimation: Approx. 10 active users – $\tau = 4$ seconds



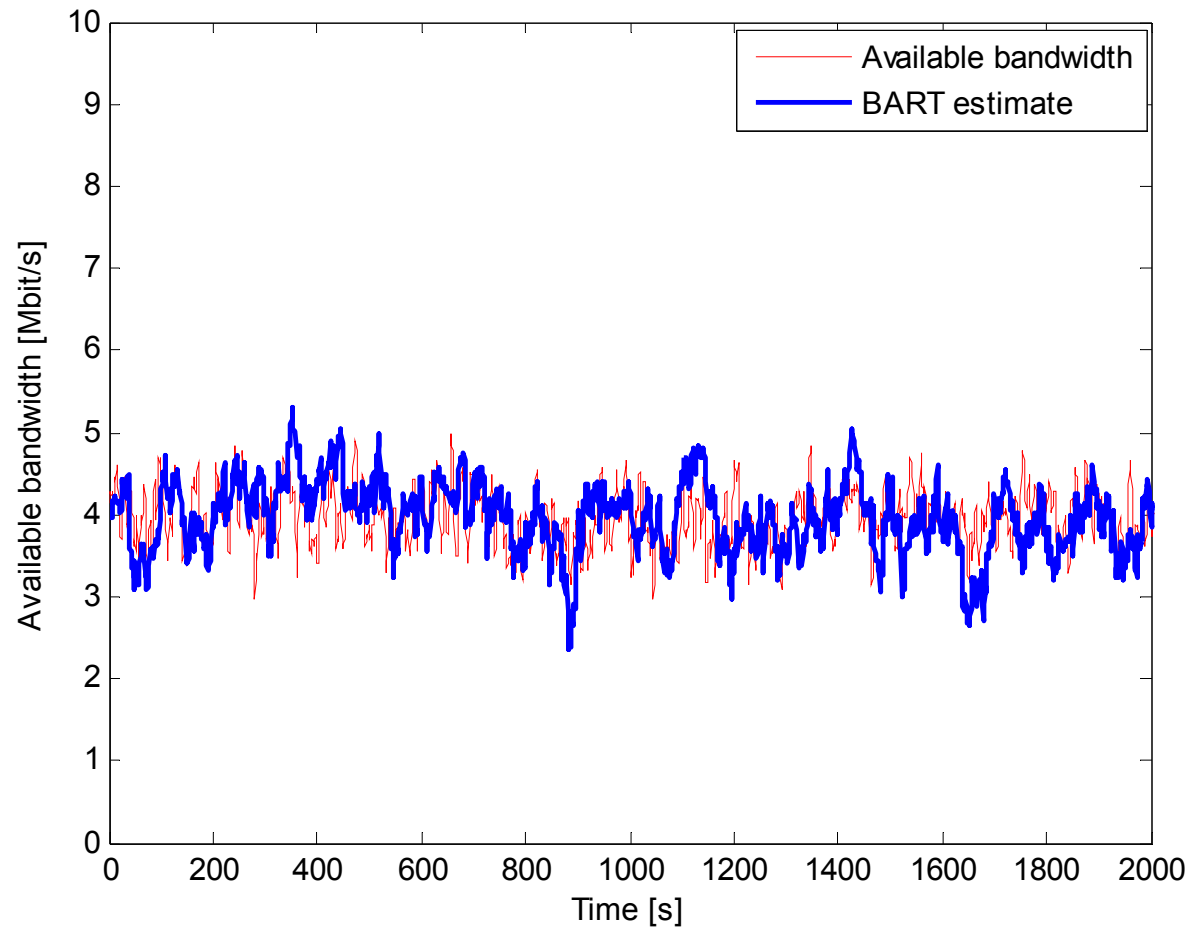
Experiments

- BART estimation: Approx. 10 active users – $\tau = 32$ seconds



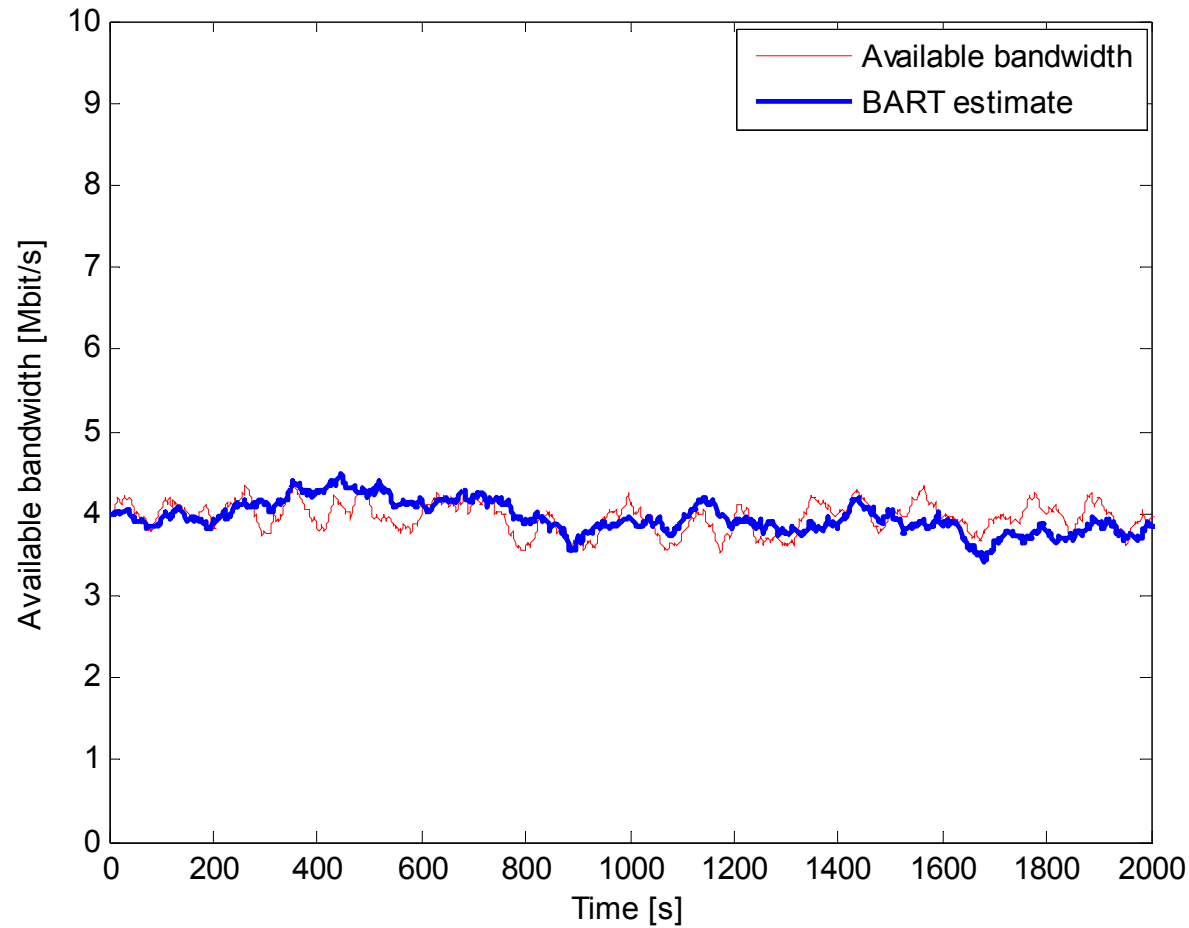
Experiments

- BART estimation: Approx. 100 active users – $\tau = 4$ seconds



Experiments

- BART estimation: Approx. 100 active users – $\tau = 32$ seconds



Summary

- Measuring available bandwidth
 - What is available bandwidth?
 - Why measuring available bandwidth?
- BART – Bandwidth Available in Real-Time
 - Active probing
 - Kalman filtering
 - The process noise covariance matrix Q
- Experiments
 - Tuning Q for optimizing the tracking for desired temporal characteristics
 - Focus on variability – $Q(2,2)$ is the crucial element
- Conclusion
 - It is feasible to tune BART for enhanced tracking performance