

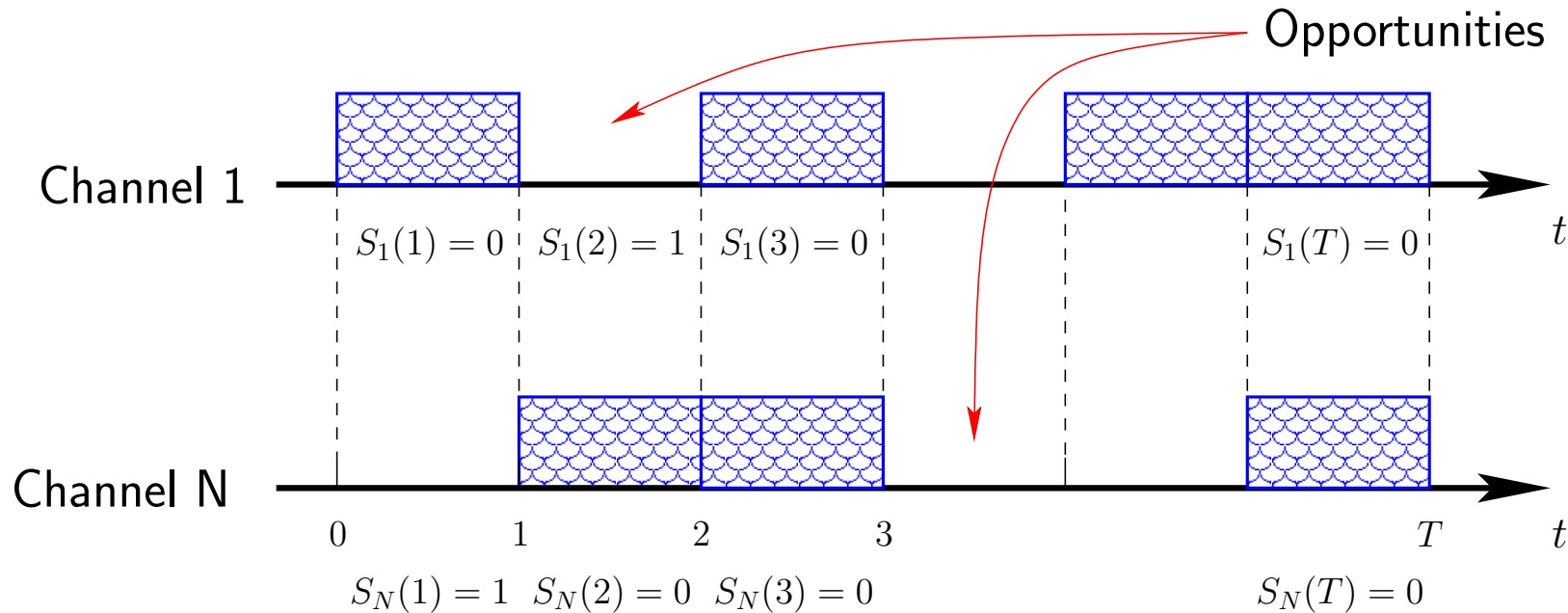
Channel Probing for Opportunistic Access with Multi-channel Sensing

Keqin Liu, Qing Zhao

Department of Electrical and Computer Engineering
University of California, Davis, CA 95616

Supported by NSF and ARL-CTA.

Multi-Channel Opportunistic Access

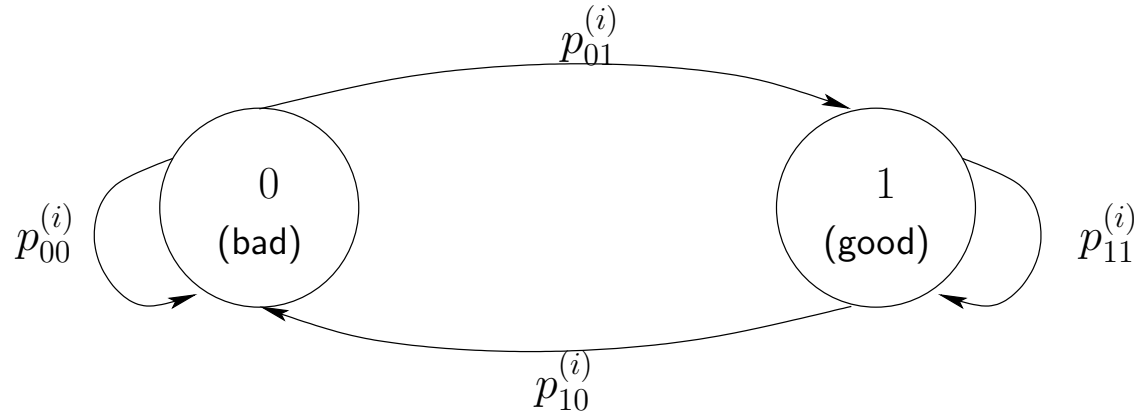


- ▶ Opportunistic Access: adapt to time-varying channel state.
- ▶ Channel State: “good (1)” or “bad (0)”
- ▶ Applications: Cognitive Radios, Downlink Scheduling in cellular network, Opportunistic transmission, Jamming/Anti-jamming
- ▶ Limited Sensing: can only sense and access K out of N channels in each slot.

Which channels to sense in each slot?

Gilbert-Elliot Channel Model

- ▶ N independent Gilbert-Elliot channels with rate B_i ($i = 1, \dots, N$).



- ▶ **Sensing Policy** π_s

- Choose the set $I(t)$ of K channels to sense in each slot t .

- ▶ **Immediate Reward**

- If sensed channel i is idle, B_i units of reward is accrued
 - If a sensed channel is busy, no reward; wait until the next slot
 - $R(t) = \sum_{i \in I(t)} S_i(t) B_i$

- ▶ **Objective:** Maximize the expected long-run reward

⁰E.N. Gilbert, "Capacity of burst-noise channels," Bell Syst. Tech. J., vol. 39, pp. 1253-1265, Sept. 1960.

Performance Measures of Long-run Reward

Two performance measures

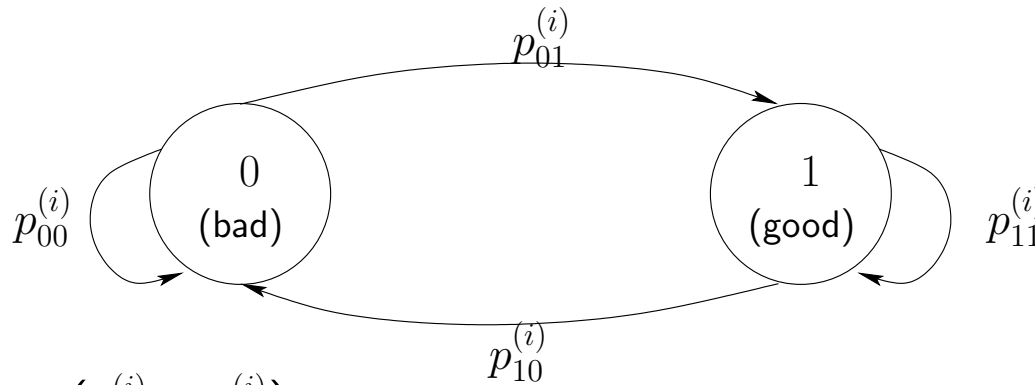
- ▶ The expected total discounted reward over an infinite horizon

$$\mathbb{E}[\sum_{t=1}^{\infty} \beta^{t-1} R(t)]$$

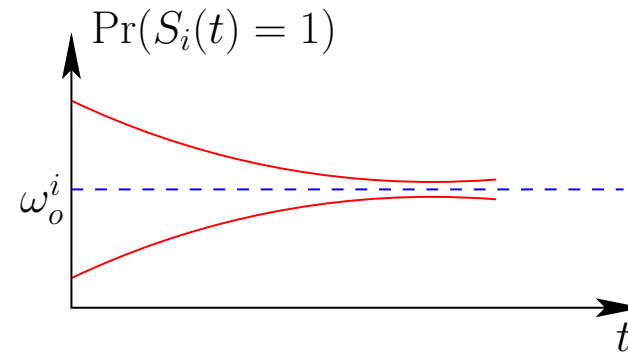
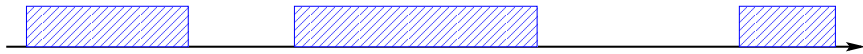
- ▶ The expected average reward over an infinite horizon

$$\mathbb{E}[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T R(t)]$$

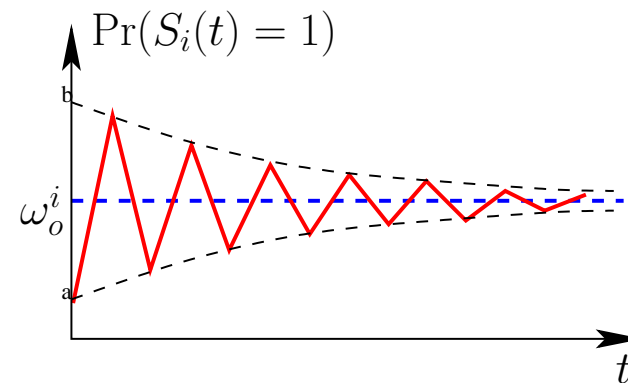
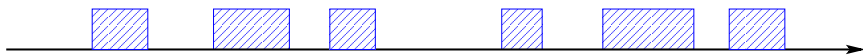
Positive Memory v.s. Negative Memory



► Positive memory ($p_{11}^{(i)} \geq p_{01}^{(i)}$)



► Negative memory ($p_{11}^{(i)} < p_{01}^{(i)}$)



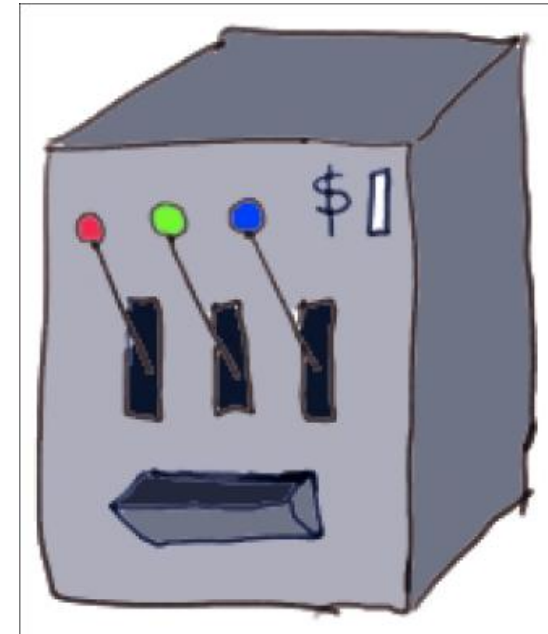
Outline

- ▶ A restless multi-armed bandit formulation
- ▶ Indexability and Whittle's index policy: discounted reward
- ▶ Indexability and Whittle's index policy: average reward
- ▶ Whittle's index policy for stochastically identical channels
- ▶ Conclusion

Multi-Armed Bandit

Multi-armed Bandit Process

- ▶ A bandit with N independent arms
- ▶ Fully observable states of all arms $\{Z_i(t)\}$
- ▶ Can activate one arm at each time
- ▶ Activate arm i and get reward $R_i(Z_i(t))$
- ▶ Active arm changes state (Markovian)
- ▶ Passive arms are frozen.



Objective: Decide which arm to activate in each slot for max long-run reward.

Complexity and Index Policy

A 2-arm Example:

State of arm 1: $\{A, B\}$

State of arm 2: $\{\alpha, \gamma\}$

- ▶ Policy: $\{A, \alpha\} \rightarrow 1$, $\{A, \gamma\} \rightarrow 2$, $\{B, \alpha\} \rightarrow 2$ and $\{B, \gamma\} \rightarrow 1$
- ▶ Complexity: Exponential with N

Low-complexity Policy: Index Policy

- Compute an index for the state of each arm
- Activate the one with the largest index
- ▶ Index assignment: $A \rightarrow X_A$, $B \rightarrow X_B$, $\alpha \rightarrow X_\alpha$, $\gamma \rightarrow X_\gamma$
- ▶ Current state $(A, \alpha) \rightarrow$ Activate arm 1 iff $X_A > X_\alpha$

Advantage: To compute index, only need to look at one arm

- ▶ Complexity: Linear with N
- ▶ Optimal Policy for Multi-armed Bandit: [Gittins' index policy \(1979\)](#)

⁰J.C.Gittins, "Bandit Processes and Dynamic Allocation Indices," in *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol.41, No.2 (1979), 148-177.

Restless Multi-Armed Bandit

Restless Multi-armed Bandit Process (Whittle'88)

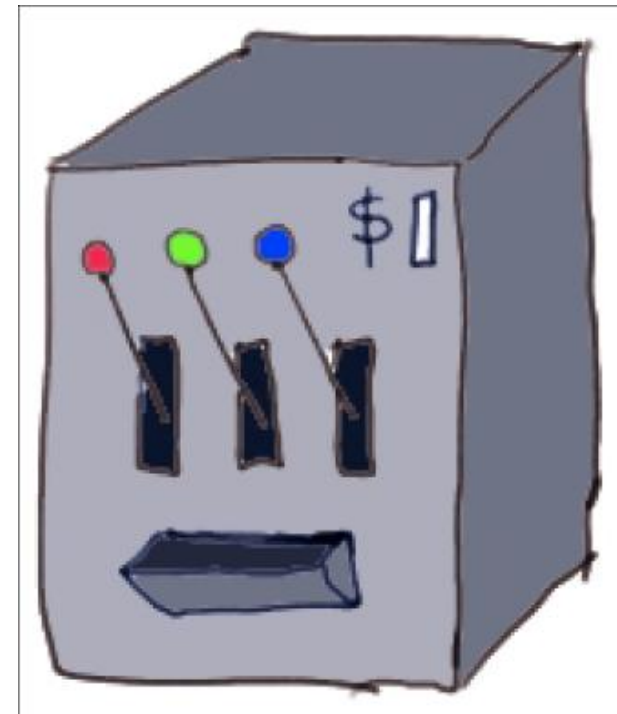
- ▶ Activate K arms
- ▶ Passive arms also change states.

Structure of Optimal Policy

- ▶ Not yet found.

Complexity

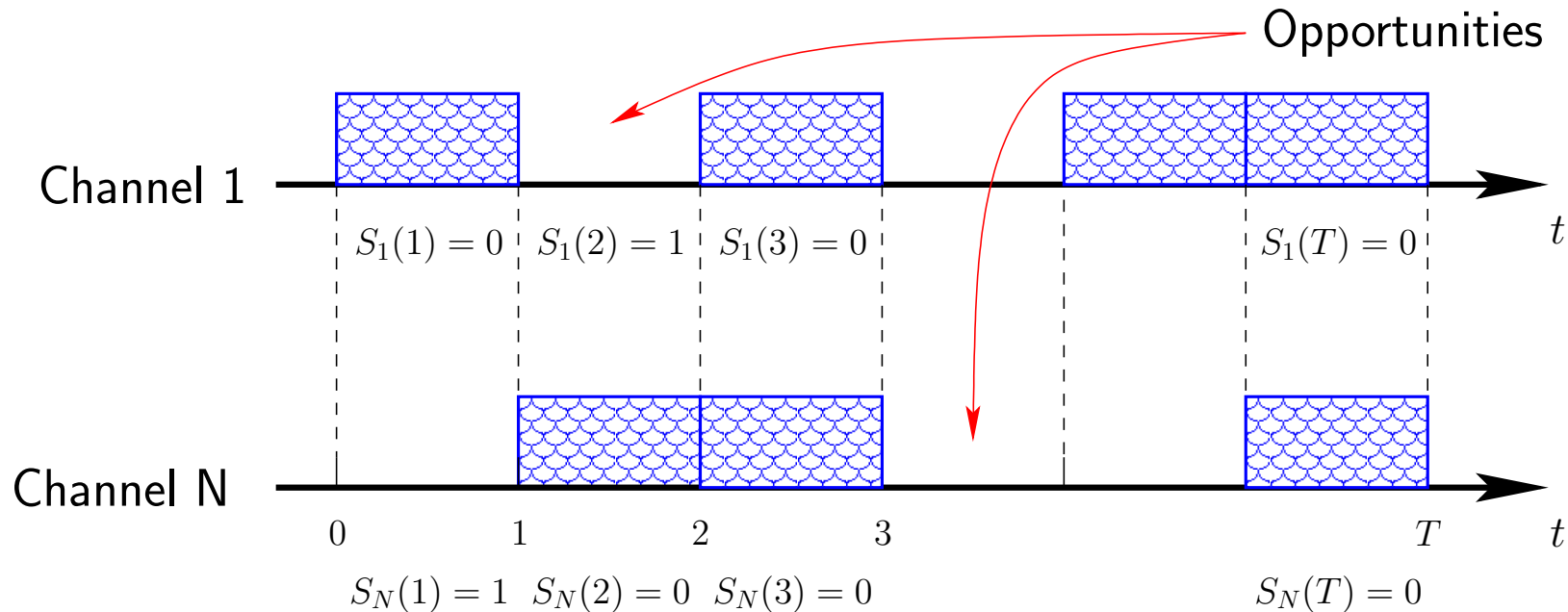
- ▶ PSPACE-hard.



⁰P. Whittle, "Restless bandits: Activity allocation in a changing world", in *Journal of Applied Probability*, Volume 25, 1988.

⁰C. H. Papadimitriou and J. N. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," in *Mathematics of Operations Research*, Vol. 24, No. 2, May 1999, pp. 293-305.

Restless Multi-armed Bandit Formulation



- ▶ Each channel is considered as an arm.
- ▶ If channel i is sensed, then it is “activated”.
- ▶ The channel state $\{S_1, \dots, S_N\}$ is not fully observable
 \Rightarrow Cannot use the channel state as the arm state

Restless Multi-armed Bandit Formulation

- ▶ The state of each arm should be the observation history of that channel.
- ▶ Sufficient statistic: the **a posterior** distribution (belief vector) $\Omega(t)$ that exploits the **entire observation history**.

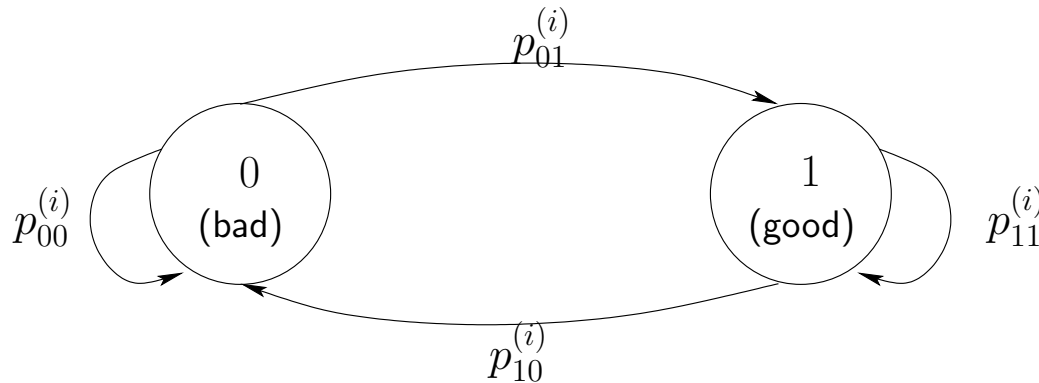
$$\Omega(t) = [\omega_1(t), \dots, \omega_N(t)]$$

$$\omega_i(t) = \Pr[\text{channel } i \text{ is idle in slot } t \mid \underbrace{O(1), \dots, O(t-1)}_{\text{observations}}]$$

- ▶ The state of arm i in slot t is $\omega_i(t)$ **uncountable state space**.
- ▶ The expected immediate reward obtained when activate arm i is $\omega_i(t)B_i$.

Markovian Transition of Belief

- ▶ The belief vector transits according to Markov processes.



- ▶ If channel i is activated in slot t :

$$\omega_i(t+1) = \begin{cases} p_{11}^{(i)}, & \text{if } O_i(t) = 1 \\ p_{01}^{(i)}, & \text{if } O_i(t) = 0 \end{cases} .$$

- ▶ If channel i is made passive in slot t :

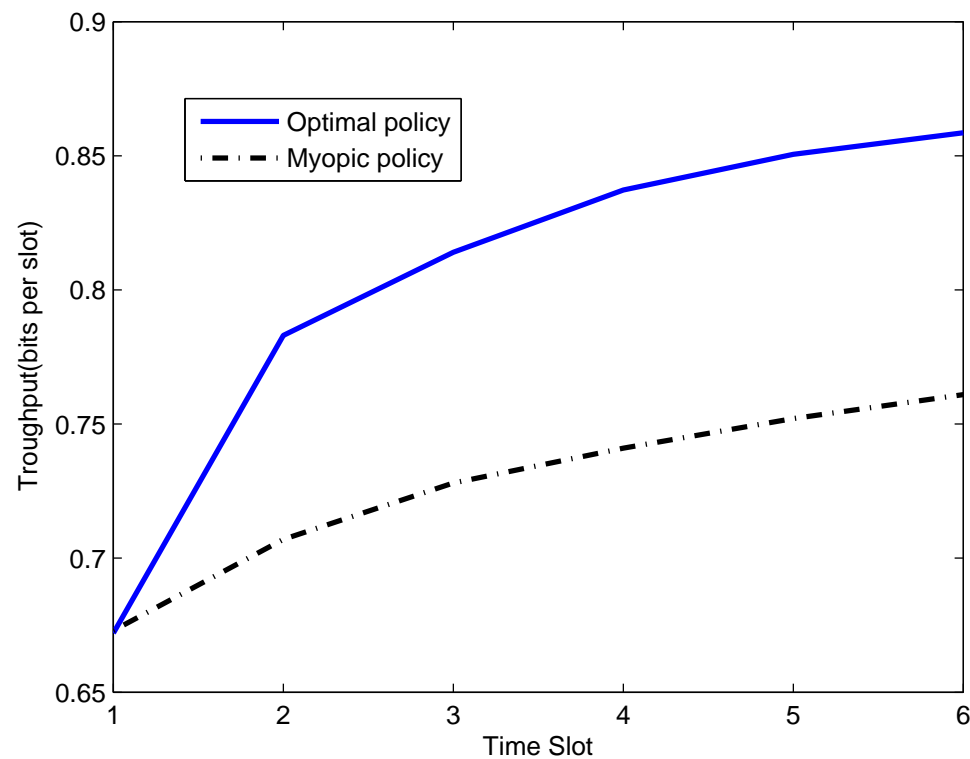
$$\omega_i(t+1) = \omega_i(t)p_{11}^{(i)} + (1 - \omega_i(t))p_{01}^{(i)} .$$

Index Policies

Are there simple index policies with good performance?

Myopic policy: maximize expected immediate reward

- ▶ Index of channel i : $W_i(t) = \mathbb{E}[R_i(t)] = \omega_i(t)B_i$



Whittle's Index Policy

Whittle's Index

- ▶ Subsidy for passivity: provide a subsidy m when the arm is made passive.
- ▶ Whittle's index: the subsidy m that makes active and passive actions equally attractive at the current state.

Performance

- Optimal under relaxed constraint on the average number of active arms.
- Asymptotically optimal ($N \rightarrow \infty$ w. $\frac{K}{N}$ fixed) under certain conditions.
- Near optimal performance observed from extensive numerical examples.

Difficulties

- Existence of Whittle's index (indexability) is often difficult to establish.
- High complexity to compute index for uncountable states of each arm.

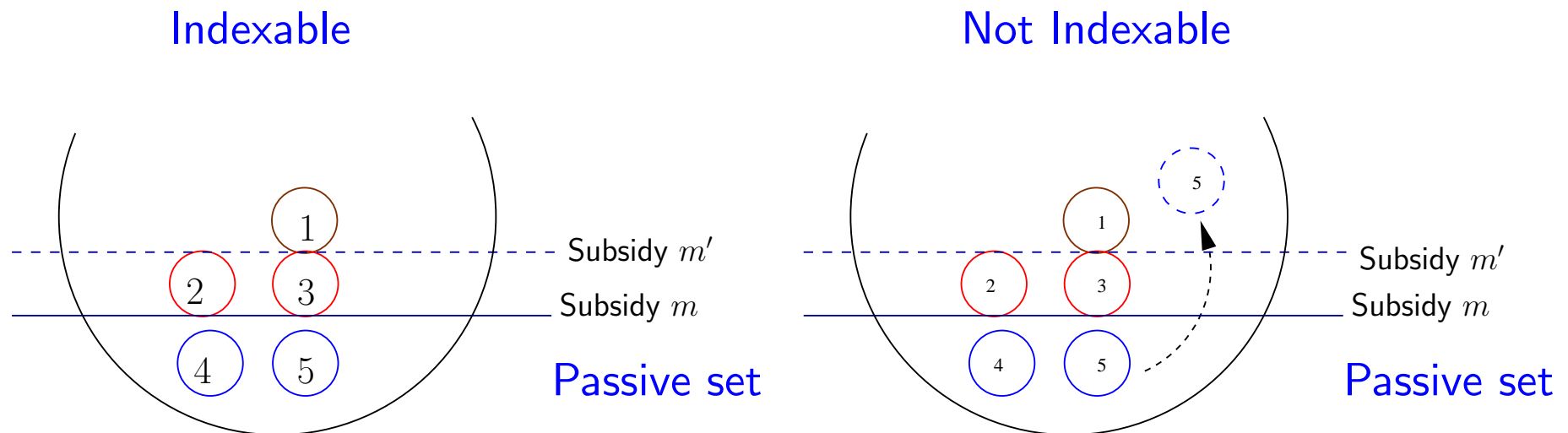
⁰P. Whittle, "Restless bandits: Activity allocation in a changing world", in *Journal of Applied Probability*, Volume 25, 1988.

⁰Richard R. Weber; Gideon Weiss, "On an Index Policy for Restless Bandits," in *Journal of Applied Probability*, Vol.27, No.3. (Sep.. 1990), pp. 637-648.

Indexability

Indexability

- ▶ $D(m)$: the set of states for which the arm should be made passive under subsidy m .
- ▶ Indexability: $D(m)$ increases monotonically from \emptyset to \mathbb{Z} as m increases from $-\infty$ to ∞ .



Single-armed Bandit Process with Subsidy

- ▶ To establish the indexability and calculate Whittle's index
 - ⇒ Sufficient to focus on the single-armed bandit process with subsidy m
- ▶ Without loss of generality, set $B = 1$.
- ▶ Policy $\pi : \omega(t) \rightarrow \{\text{Active}, \text{Passive}\}$

$R(t):$ $\omega(1)$ m m $\omega(4)$ m
 _____ | Active | Passive | Passive | Active | Passive | _____

belief $\omega(t):$ $\omega(1)$ $\omega(2)$ $\omega(3)$ $\omega(4)$ $\omega(5)$

- ▶ The objective is to maximizing the total discounted reward

$$\max_{\pi} \{ \mathbb{E}_{\pi} [\sum_{t=1}^{\infty} \beta^{t-1} R(t) | \omega(1)] \}$$

where β ($0 \leq \beta < 1$) is a discount factor.

Value Function for Discounted Reward Criterion

- ▶ **Value Function $V_m(\omega)$** : The maximum expected total discounted reward under subsidy m starting from belief ω
- ▶ **$V_m(\omega; \text{active})$** : the total discounted reward by being active first then followed by the optimal policy starting from belief ω

$$V_m(\omega; \text{active}) = \omega + \beta(\omega V_m(p_{11}) + (1 - \omega)V_m(p_{01})) \quad (\text{Linear})$$

- ▶ **$V_m(\omega; \text{passive})$** : the total discounted reward by being passive first then followed by the optimal policy starting from belief ω

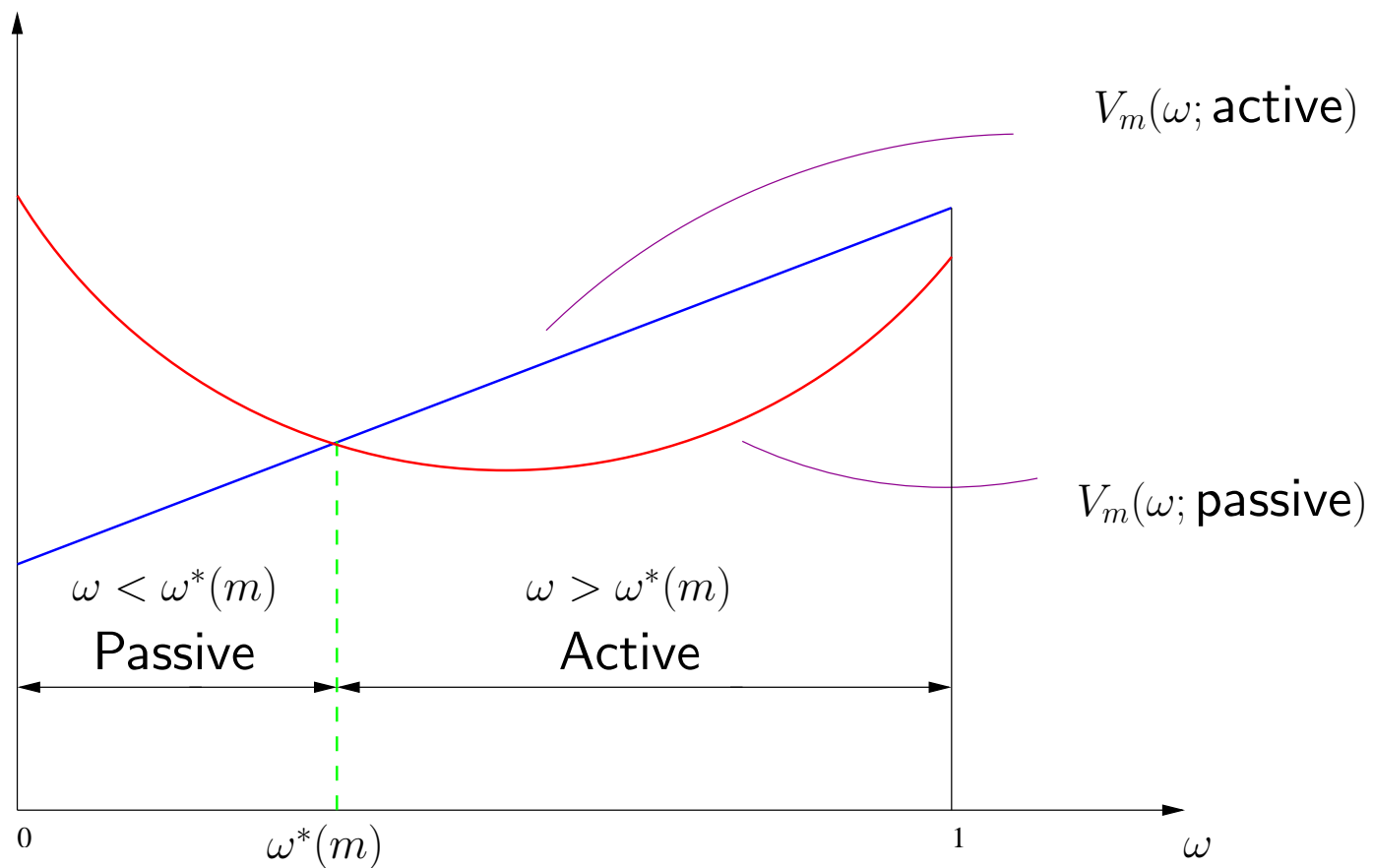
$$V_m(\omega; \text{passive}) = m + \beta V_m(\mathcal{T}(\omega)) \quad (\text{Convex})$$

where $\mathcal{T}(\omega) = \omega p_{11} + (1 - \omega)p_{01}$ denotes the one-step belief update when passive.

$$V_m(\omega) = \max\{V_m(\omega; \text{active}), V_m(\omega; \text{passive})\}$$

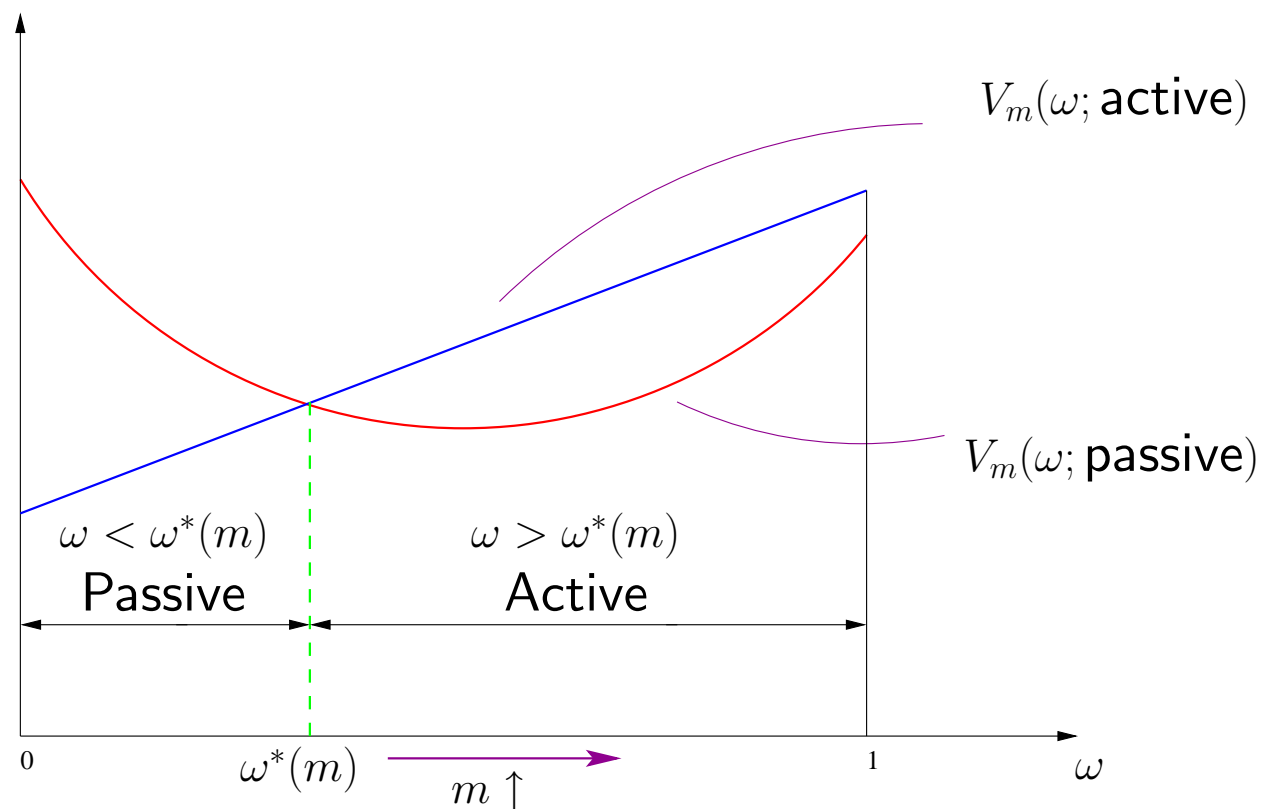
The Optimality of Threshold Policy and Indexability

- ▶ The optimal policy for the single-armed bandit is a threshold policy



The Optimality of Threshold Policy and Indexability

- ▶ The optimal policy for the single-armed bandit is a threshold policy



- ▶ Indexability: $\omega^*(m) \uparrow$ with m

Indexability

To show $\omega^*(m)$ is increasing with m

- ▶ need to show $\omega^*(m)$ stays in the passive set under a larger subsidy

$$\frac{d(V_m(\omega; \text{passive}))}{dm} \Big|_{\omega=\omega^*(m)} \geq \frac{d(V_m(\omega; \text{active}))}{dm} \Big|_{\omega=\omega^*(m)}$$

- ▶ The derivative of value function is the total discounted time of being passive
- ▶ By analyzing the passive time, we establish the indexability.

Solve for Whittle's Index in Closed-Form

- ▶ Whittle's index $W(\omega)$ is the subsidy m that makes active and passive actions equally attractive at ω

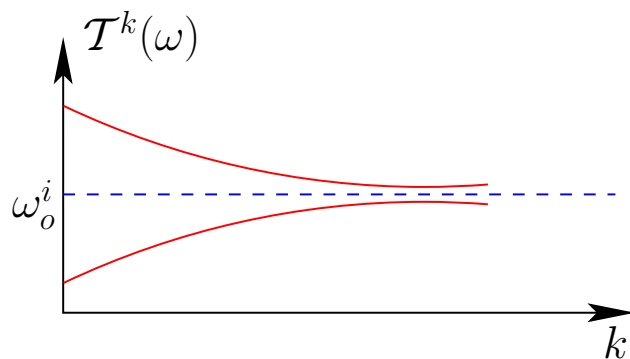
$$V_m(\omega; \text{active}) = V_m(\omega; \text{passive})$$

- ▶ Need to solve for the value functions in closed-form

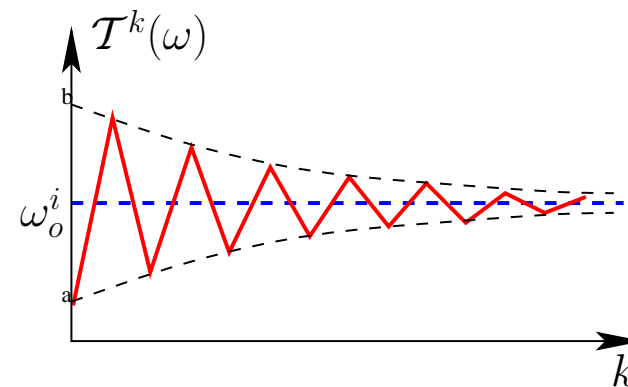
Key: Obtain $V_m(\omega; \text{active})$ and $V_m(\omega; \text{passive})$ as functions of $V_m(p_{11})$ and $V_m(p_{01})$

- ▶ $V_m(\omega; \text{active}) = \omega + \beta(\omega V_m(p_{11}) + (1 - \omega)V_m(p_{01}))$ and $V_m(\omega; \text{passive}) = m + \beta V_m(\mathcal{T}(\omega))$

Positive Memory



Negative Memory



- ▶ $V_m(p_{11})$ and $V_m(p_{01})$ can be obtained in closed-form

Whittle's Index in Closed-form

► Positive memory ($p_{11} \geq p_{01}$)

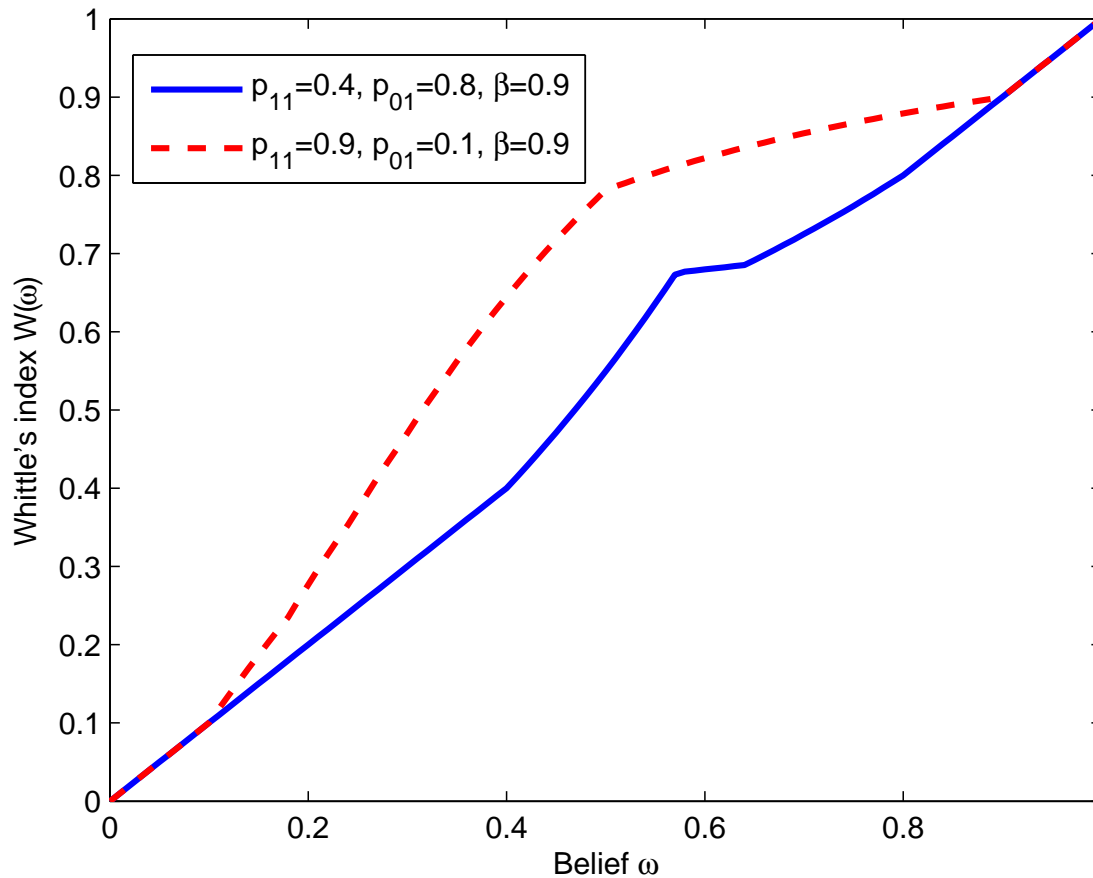
$$W(\omega) = \begin{cases} \omega B_i, & \text{if } \omega \leq p_{01} \text{ or } \omega \geq p_{11}; \\ \frac{\omega}{1-\beta p_{11}+\beta\omega} B_i, & \text{if } \omega_o \leq \omega < p_{11}; \\ \frac{\omega-\beta\mathcal{T}(\omega)+C(1-\beta)(\beta(1-\beta p_{11})-\beta(\omega-\beta\mathcal{T}(\omega)))}{1-\beta p_{11}-A(1-\beta)(\beta(1-\beta p_{11})-\beta(\omega-\beta\mathcal{T}(\omega)))} B_i, \\ \text{if } p_{01} < \omega < \omega_o; \end{cases}$$

► Negative memory ($p_{11} < p_{01}$)

$$W(\omega) = \begin{cases} \omega B_i, & \text{if } \omega \leq p_{11} \text{ or } \omega \geq p_{01}; \\ \frac{\beta p_{01}+\omega(1-\beta)}{1+\beta(p_{01}-\omega)} B_i, & \text{if } \mathcal{T}(p_{11}) \leq \omega < p_{01}; \\ \frac{(1-\beta)(1+\beta E)(\beta p_{01}+\omega(1-\beta))}{1-\beta(1-p_{01})-D(1-\beta)(\beta^2 p_{01}+\beta\omega-\beta^2\omega)} B_i, \\ \text{if } \omega_o \leq \omega < \mathcal{T}(p_{11}); \\ \frac{(1-\beta)(\beta p_{01}+\omega-\beta\mathcal{T}(\omega))-E(1-\beta)\beta(\beta\mathcal{T}(\omega)-\beta p_{01}-\omega)}{1-\beta(1-p_{01})+D(1-\beta)\beta(\beta\mathcal{T}(\omega)-\beta p_{01}-\omega)} B_i, \\ \text{if } p_{11} < \omega < \omega_o; \end{cases}$$

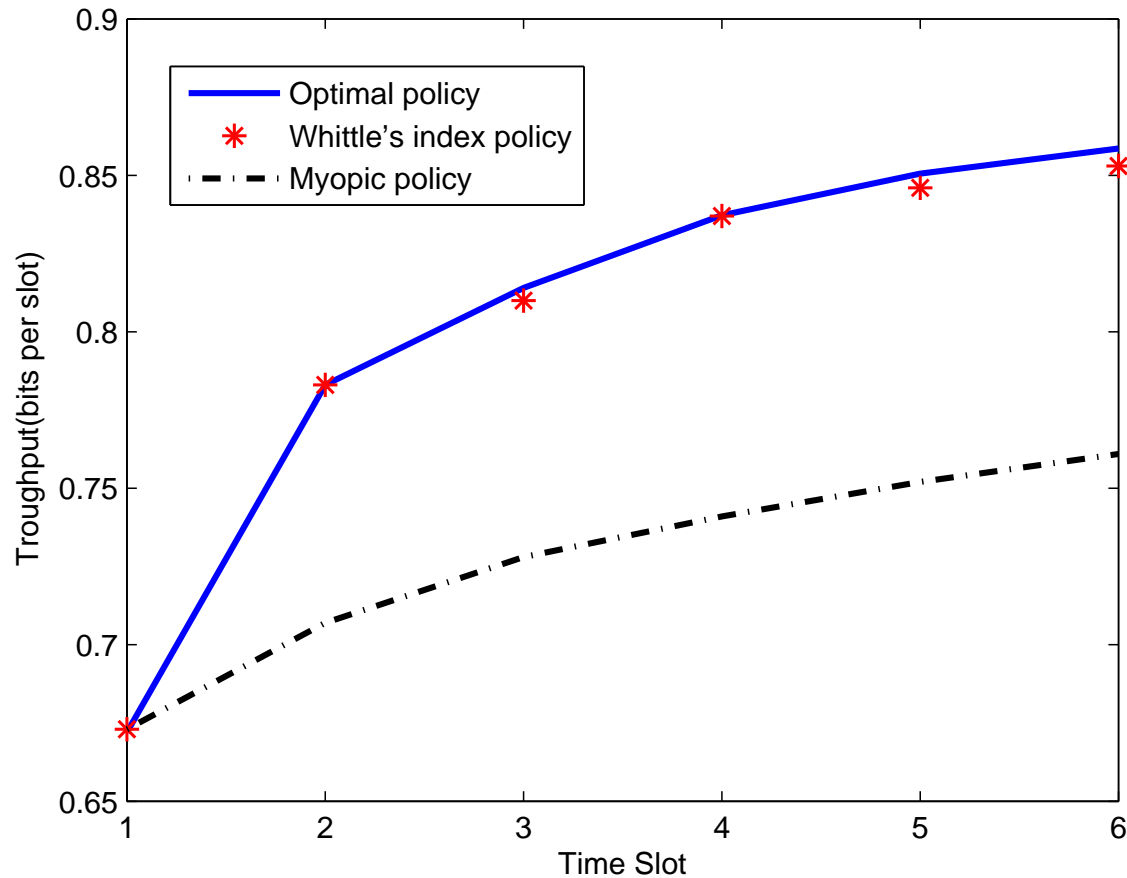
► Simple implementation of Whittle's Index Policy

Monotonicity of Whittle's index



- ▶ Whittle's index is an increasing function of the belief state.
- ▶ Whittle's index policy is equivalent to myopic policy for identical channels.

The Performance of Whittle's Index Policy



- ▶ Need a performance benchmark that can be computed with low complexity

An Upper Bound of the Optimal Performance

Upper Bound of the Optimal Performance

- ▶ Relax the sensing constraint: only require to sense K channels on average
- ▶ Inexplicit form of the upper bound by Lagrangian Multiplier Theorem

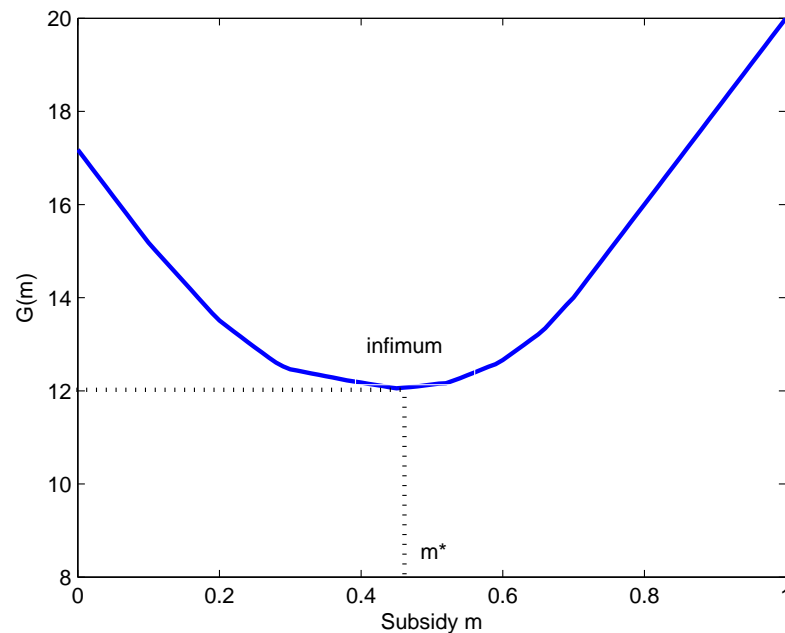
$$V(\Omega) \leq \inf_m \left\{ \underbrace{\sum_{i=1}^N V_m^i(\omega_i)}_{G(m)} - m \frac{N - K}{1 - \beta} \right\}.$$

- ▶ $V_m^i(\omega_i)$ is the value function for single arm i with subsidy m
- ▶ $G(m)$ can be obtained in closed-form
- ▶ An efficient algorithm to find the infimum over m

⁰P. Whittle, "Restless bandits: Activity allocation in a changing world", in *Journal of Applied Probability*, Volume 25, 1988.

An Upper Bound of the Optimal Performance

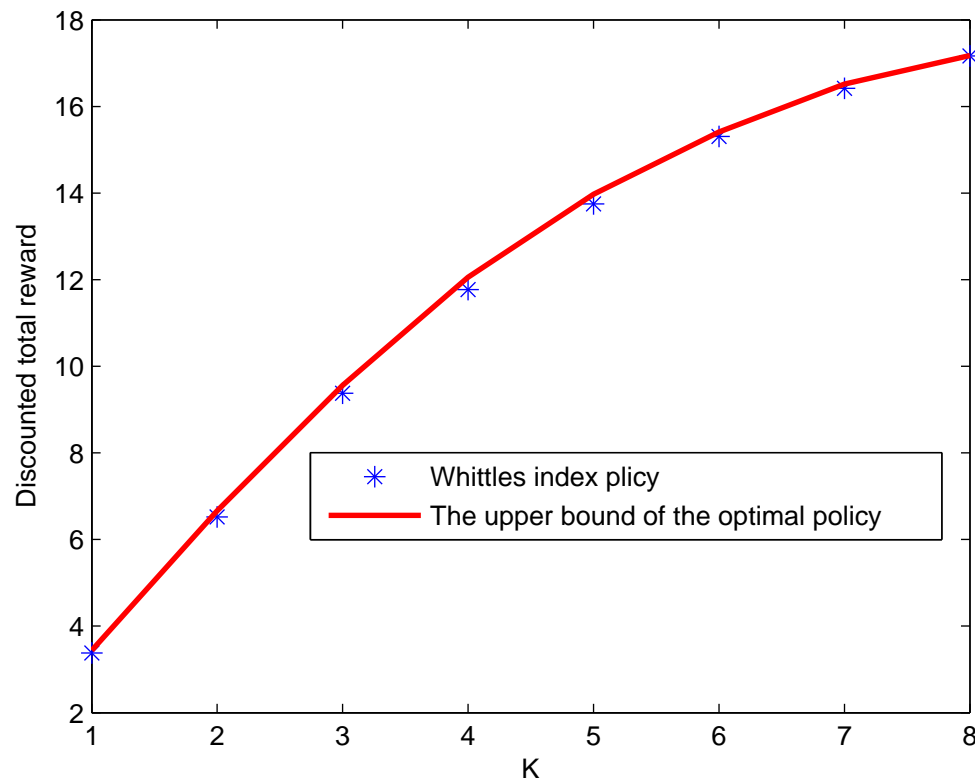
- ▶ $G(m)$ is convex in m , but may not be differentiable as in this case.



- ▶ The upper bound is achieved when the sign of $\frac{dG(m)}{dm}$ is about to change
- ▶ $G(m)$ in closed-form \implies its derivative in closed-form \implies easy search of the minimizing subsidy

The Performance of Whittle's Index Policy

- ▶ The tightness of the performance upper bound
- ▶ The near-optimal performance of Whittle's index policy



Outline

- ▶ A restless multi-armed bandit formulation
- ▶ Indexability and Whittle's index policy: discounted reward
- ▶ Indexability and Whittle's index policy: average reward
- ▶ Whittle's index policy for stochastically identical channels
- ▶ Conclusion

Average Reward Criterion

- ▶ From discounted reward to average reward (Dutta'1991)

- ▶ **Valueboundedness Condition:** The difference between two discounted value functions starting from different belief values is uniformly bounded..
 - π_β^* pointwise converges to π^* as $\beta \rightarrow 1$

 - The maximum average reward $J(\Omega) = \lim_{\beta \rightarrow 1} (1 - \beta)V_\beta(\Omega)$

⁰P. K. Dutta, "What do discounted optima converge to? A theory of discount rate asymptotics in economic models," in Journal of Economic Theory 55, pp. 6494, 1991.

Average Reward Criterion

- ▶ Value boundedness condition can be verified from the closed-form value functions for single-armed bandit
- ▶ All results for discounted reward can be extended to average reward
 - Establish indexability by examining $\lim_{\beta \rightarrow 1} \pi_{\beta}^*$
 - Whittle's index: Taking the limit of Whittle's index for discounted reward
 - Performance upperbound: Taking the limit of the upper bound for discounted reward

Whittle's Index under Average Reward Criterion

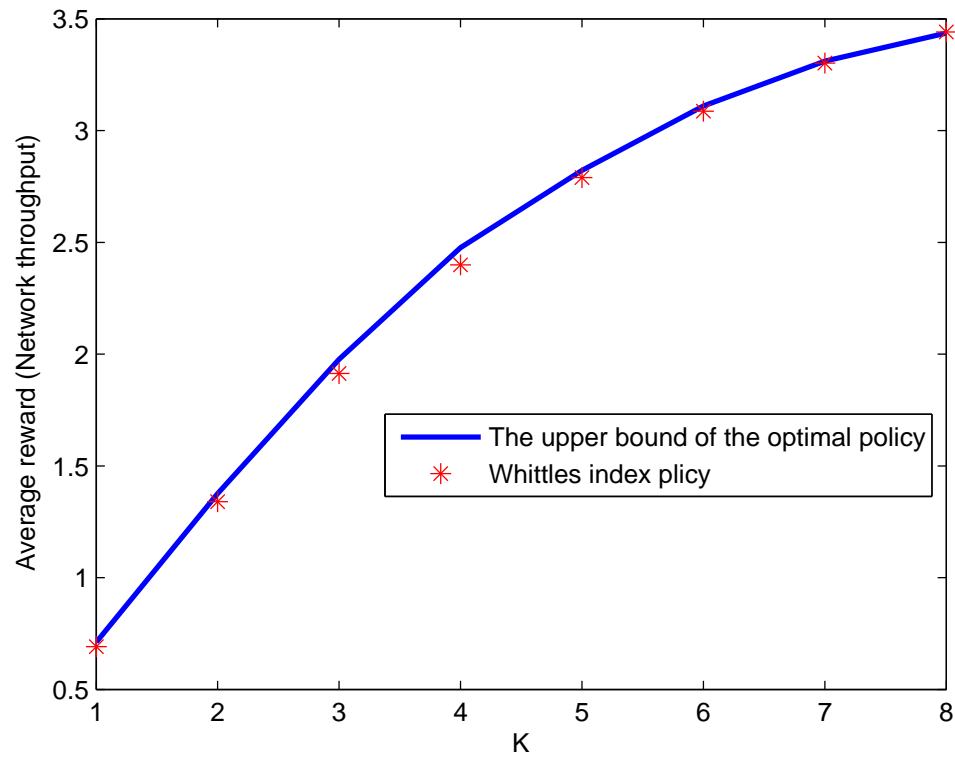
► Positive memory ($p_{11} \geq p_{01}$)

$$W(\omega) = \begin{cases} \omega B_i, & \text{if } \omega \leq p_{01} \text{ or } \omega \geq p_{11} \\ \frac{\omega}{1-p_{11}+\omega} B_i, & \text{if } \omega_o \leq \omega < p_{11} \\ \frac{(\omega - \mathcal{T}^1(\omega)) * (L+2) + \mathcal{T}^{L+1}(p_{01})}{1-p_{11} + (\omega - \mathcal{T}^1(\omega)) * (L+1) + \mathcal{T}^{L+1}(p_{01})} B_i, & \text{if } p_{01} < \omega < \omega_o \end{cases}$$

► Negative memory ($p_{11} < p_{01}$)

$$W(\omega) = \begin{cases} \omega B_i, & \text{if } \omega \leq p_{11} \text{ or } \omega \geq p_{01} \\ \frac{p_{01}}{1+p_{01}-\omega} B_i, & \text{if } \mathcal{T}^1(p_{11}) \leq \omega < p_{01} \\ \frac{p_{01}}{1+p_{01}-\mathcal{T}^1(p_{11})} B_i, & \text{if } \omega_o \leq \omega < \mathcal{T}^1(p_{11}) \\ \frac{\omega + p_{01} - \mathcal{T}^1(\omega)}{1+p_{01}-\mathcal{T}^1(p_{11}) + \mathcal{T}^1(\omega) - \omega} B_i & \text{if } p_{11} < \omega < \omega_o \end{cases}$$

Performance of Whittle's Index Policy



Outline

- ▶ A restless multi-armed bandit formulation
- ▶ Indexability and Whittle's index policy: discounted reward
- ▶ Indexability and Whittle's index policy: average reward
- ▶ Whittle's index policy for stochastically identical channels
- ▶ Conclusion

Whittle's Index Policy for stochastically identical channels

For *stochastically identical* channels, Whittle's index policy is equivalent to the myopic policy.

The Myopic Policy under Single-Channel Sensing ($K = 1$)

- ▶ A semi-universal structure (Zhao&Krishnamachari'2007)
- ▶ The optimality of the myopic policy
 - $N = 2$ (Zhao&Krishnamachari'2007)
 - $N > 2$ when each channel has positive memory (Ahmad&etal'2008)
- ▶ Scaling behavior of throughput with respect to N
 - Throughput saturates quickly as N increases (Liu&Zhao'2008)

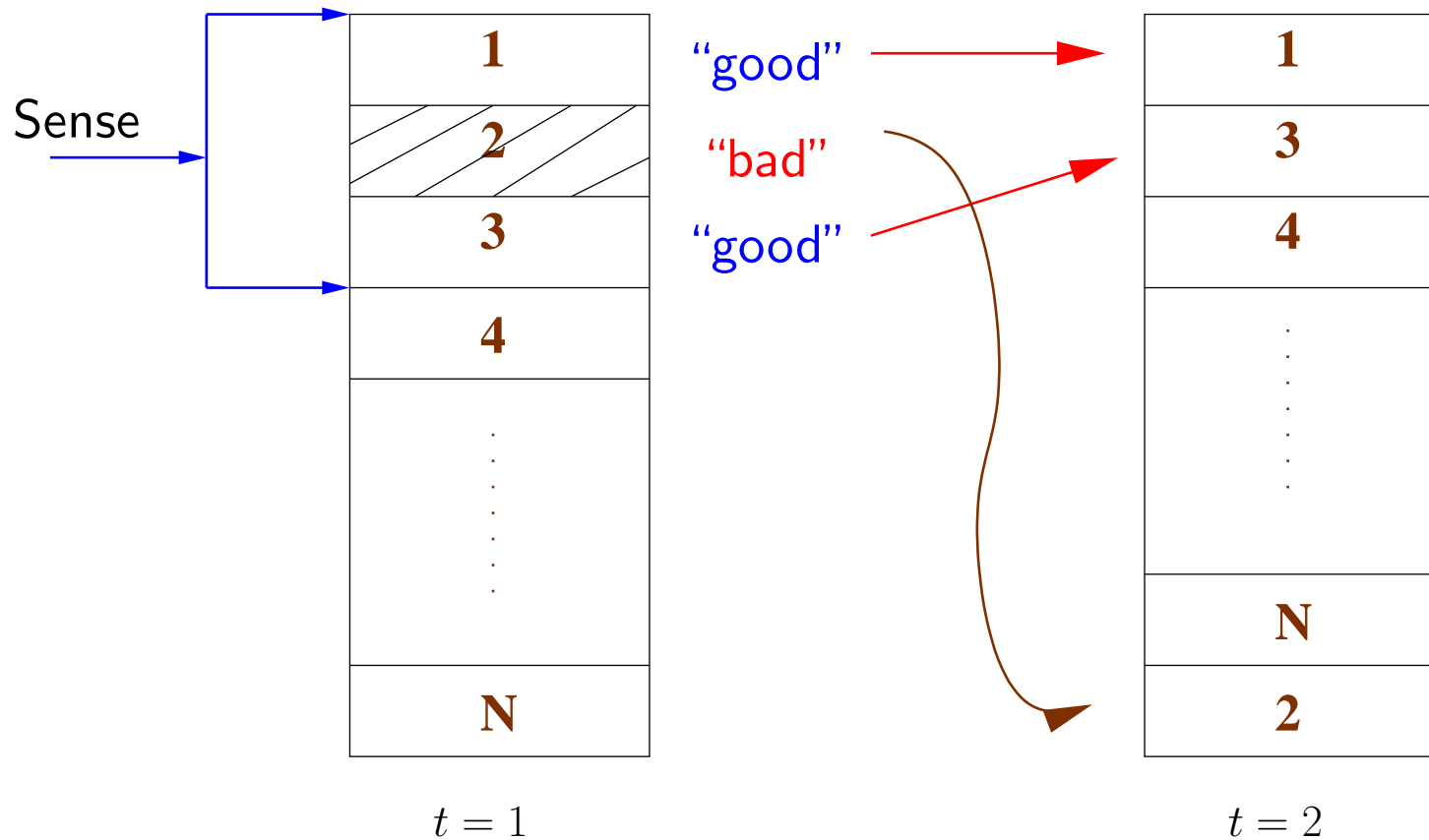
^oQ. Zhao and B. Krishnamachari, "Structure and Optimality of Myopic Sensing for Opportunistic Spectrum Access," CogNet 2007.

^oS.H. Ahmad, M. Liu, T. Javidi, Q. Zhao and B. Krishnamachari, "Optimality of Myopic Sensing in Multi-Channel Opportunistic Access," submitted to IEEE Transactions on Information Theory, May, 2008.

^oK. Liu and Q. Zhao, "Link Throughput of Multi-Channel Opportunistic Access with Limited Sensing," ICASSP, 2008.

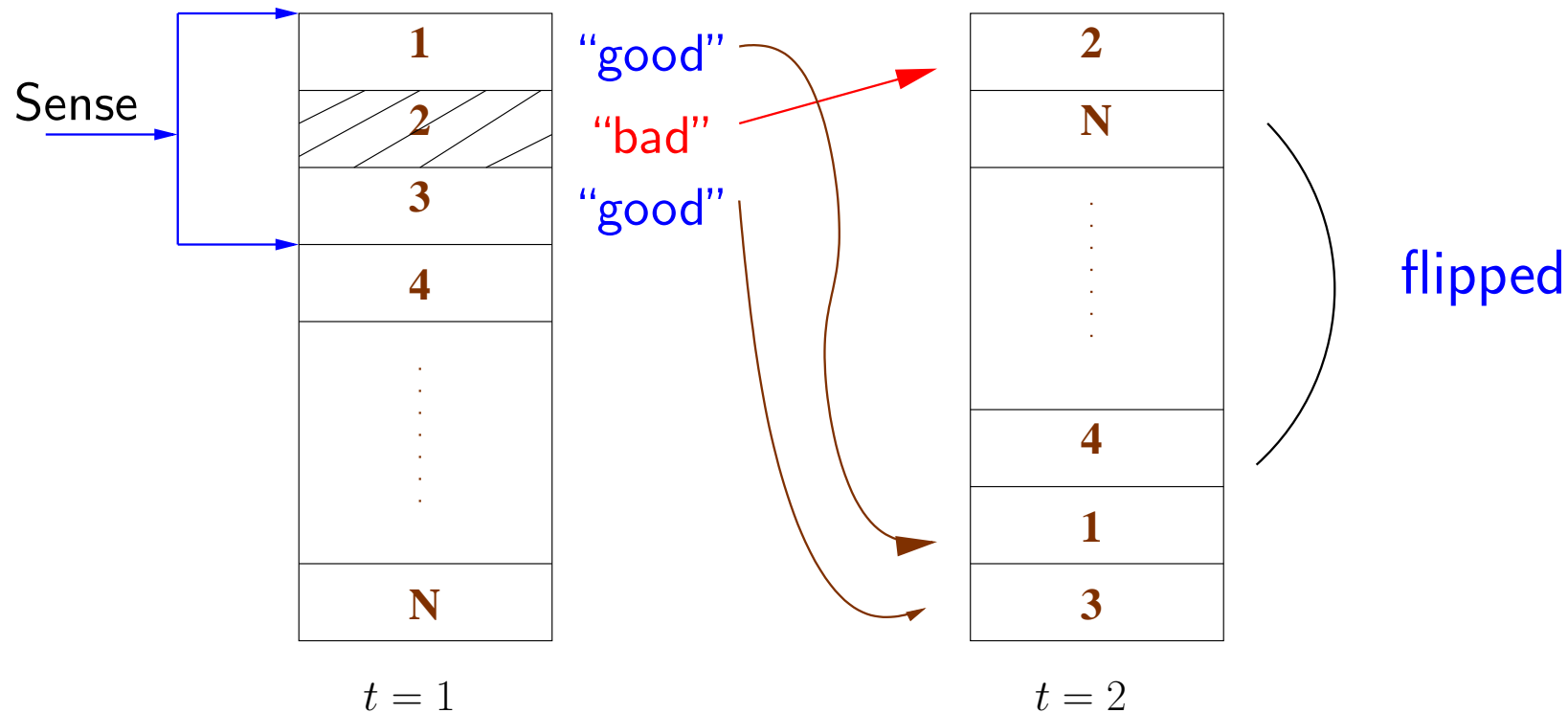
Structure of Whittle's Index Policy for Identical Channels: Positive Memory

- Stay with idle channels and leave busy ones to the end of the queue.



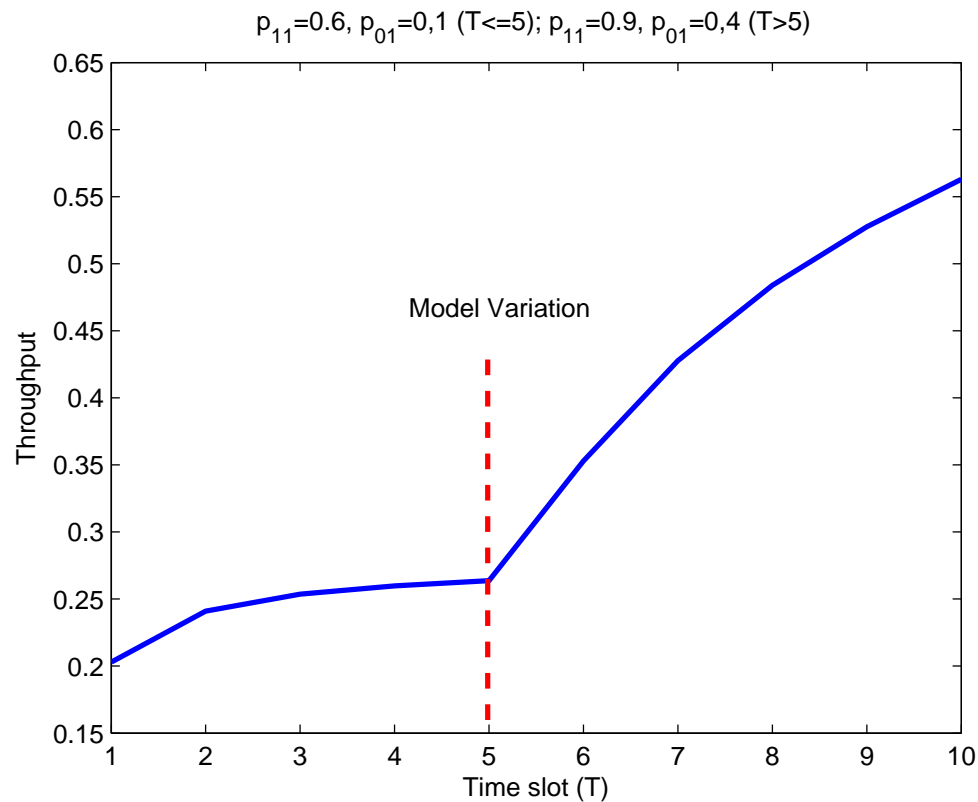
Structure of Whittle's Index Policy for Identical Channels: Negative Memory

- ▶ Stay with busy channels and leave idle ones to the end of the queue.
- ▶ Reverse the order of unobserved channels.



Robustness of Whittle's Index Policy

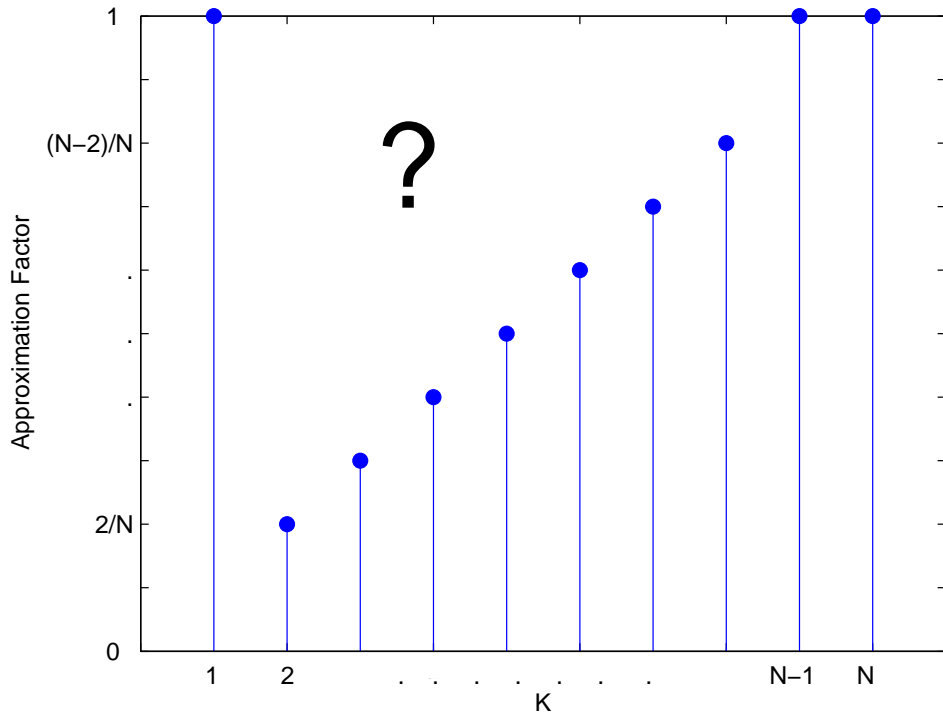
- ▶ No need to know the transition probabilities except the order of p_{11} and p_{01} .
- ▶ Automatically tracks model variations.



Approximation Factor of Whittle's Index Policy

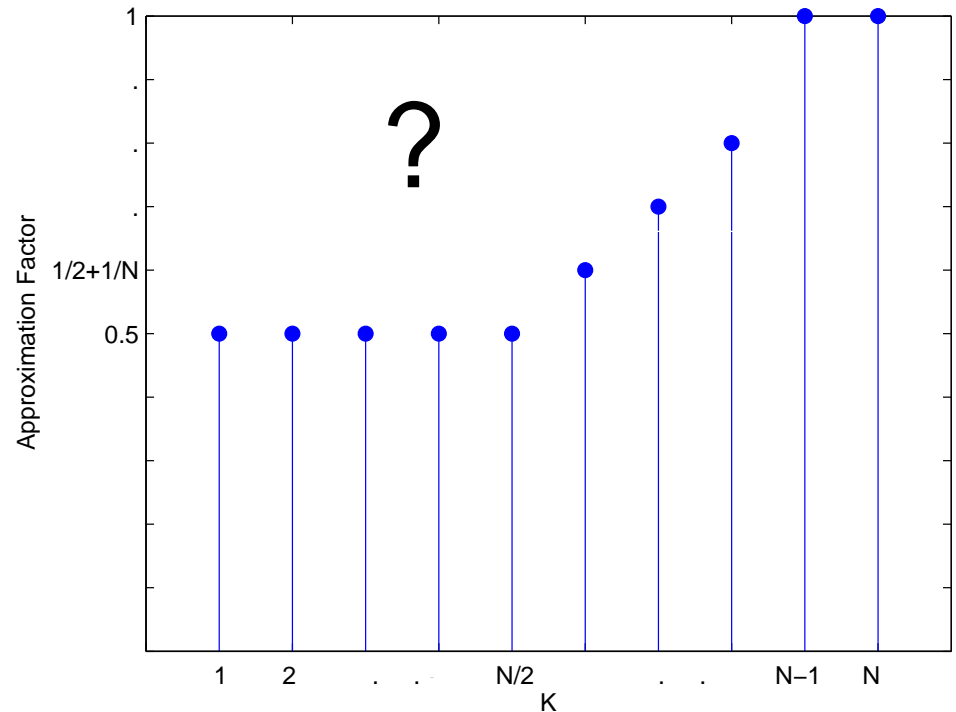
► Approximation factor $\eta = \frac{\text{Performance of myopic policy}}{\text{Optimal performance}}$

Positive Memory



$$\begin{cases} \eta = 1, & \text{for } K = 1, N - 1, N \\ \eta \geq \frac{K}{N}, & \text{o.w.} \end{cases}$$

Negative Memory



$$\begin{cases} \eta = 1, & \text{for } K = N - 1, N \\ \eta \geq \max\{\frac{1}{2}, \frac{K}{N}\}, & \text{o.w.} \end{cases}$$

Conclusion

- ▶ A restless multi-armed bandit formulation

- ▶ Indexability and Whittle's index policy: discounted and average reward
 - Establish the indexability of the system
 - Obtain Whittle's index in closed-form
 - An easily computable upper bound of the optimal performance

- ▶ Whittle's index policy for stochastically identical channels
 - A semi-universal structure: simple and robust to model mismatch and variations
 - Approximation factor of the performance