

# OPTIMAL RECONSTRUCTION OF GAUSS MARKOV FIELD IN LARGE SENSOR NETWORKS

*Min Dong and Lang Tong<sup>†</sup>*

School of Electrical and Computer Engineering  
Cornell University, Ithaca, NY 14853  
{mdong, ltong}@ece.cornell.edu

*Brian M. Sadler*

Army Research Laboratory  
Adelphi, MD 20783  
bsadler@arl.army.mil

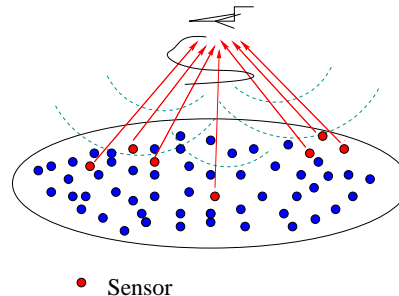
## ABSTRACT

We consider the problem of reconstructing a one-dimensional Gauss Markov field measured by a large-scale sensor network. Two data retrieval strategies are considered: the scheduling that collects data from equally spaced sensors locations and random access. Assuming the sensors in the field form a Poisson field with density  $\rho$ , we examine the reconstruction performance of the signal field based on the data retrieved under the two strategies. Our comparison shows that, the performance under the optimal scheduling is sensitive to the outage probability  $P_{\text{out}}$  of sensors in a given region. If  $P_{\text{out}}$  is large than the threshold, the performance of scheduling suffers from missing data samples, and simple random access outperforms optimal scheduling.

## 1. INTRODUCTION

Sensor network with mobile access points (SENMA), shown in Fig. 1, is proposed in [1] as a reachback network architecture for large-scale low power sensor networks. In SENMA, Sensors are low-powered operating nodes in the field that perform simple functions such as sensing/measurements. The access points are a few more powerful mobile nodes responsible for collecting data from sensors and handling sophisticated processing tasks. One type of applications in SENMA is to provide measurements of some phenomena of interest in a field, such as temperature. At a prearranged time, sensors take local measurements forming a snapshot of the field information. When the access point is ready, they transmit their packets back to the access point through a common wireless channel according to a specified medium access control (MAC) protocol. Based on received data samples, the access point reconstructs the original signal field.

In sensor networks, medium access control (MAC) governs the phase of data retrieval from a sensor field to the access point. In a dense sensor network, data among sensors are highly correlated. Every packet received at the access point provides a (noisy) data sample of the signal field at the sensor location where the packet was generated from. Intuitively, collecting data from uniformly spaced locations is better than collecting data from a concentrated sub-area because highly correlated measurements provide less information about the source. In the ideal case when there exists a sensor at any location, we have shown in [2] that, the optimal scheduling scheme which extracts data from uniform



**Fig. 1:** a 1-D reachback sensor network with a mobile access point

spaced locations provides higher efficiency on reconstruction than the ALOHA-like random access in the sense that the reconstruction distortion decay faster than that of the latter one. In terms of reconstruction performance, we have also shown that the optimal scheduling always gives better performance than random access, although the performance gain depends on the level of measurement SNR: for the high SNR regime, the performance gap is substantial; at low SNR, however, random access is shown to result in little performance loss.

In practice, however, due to finite sensor populations, there always exists areas void of sensors. The scheduling scheme in this case may not find sensors at a desired location, even at a small neighborhood of that location. Hence, it loses a data sample from this area. In such situation, scheduling may have performance limitation because of missing data samples. In contrast, such problem does not exist for the random access. How the reconstruction performance is affected by different data retrieval strategies in this case needs careful examination. To this end, we consider the performance of reconstructing a one-dimensional random signal field measured by a large-scale sensor network using different data retrieval strategies. The relative performance is shown to depend on the value of outage probability  $P_{\text{out}}$  of sensors in a given region. Our comparison shows that, although the scheduling is designed to collect data in the field using an optimal data retrieval pattern, its performance is sensitive to the outage probability  $P_{\text{out}}$  of sensors in a given region. If  $P_{\text{out}}$  is large than a threshold, the performance of scheduling suffers from missing data samples, and simple random access outperforms optimal scheduling.

The MAC design problem for sensor networks has attracted a growing interest. Many MAC protocols have been proposed aiming to the specific requirements for both ad hoc sensor networks [3–5] and reachback sensor networks [6, 7]. Among these protocols, system throughput and energy efficiency are the major

<sup>†</sup>This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, and Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011.

considerations. For applications such as estimating the signal field using a sensor network, how these MAC schemes actually affect the estimation and reconstruction performance has not been addressed. Perhaps the most relevant work is [8], where the problem of random sampling of a dynamic system is discussed, and sequential estimation of a dynamic source is considered using random sampling in time.

## 2. PROBLEM STATEMENT

### 2.1. Source Model

Consider a one-dimensional field of length  $D$  denoted by  $\mathcal{A} = [0, D]$ . Let  $S_t(x)$  ( $x \in \mathcal{A}$ ) be the source of interest in  $\mathcal{A}$  at time  $t$ . We assume that the spatial dynamic of  $S_t(x)$  is a one-dimensional homogeneous Gaussian random field governed by the following linear stochastic differential equation:

$$dS_t(x) = -fS_t(x)dx + \sigma dW_t(x) \quad (1)$$

where  $f > 0$  and  $\sigma$  are known; The process  $W_t(x)$  is a standard Brownian motion, and the source signal  $S_t(x)$  is the stationary solution of (1) with  $S_t(x) \sim \mathcal{N}(0, \frac{\sigma^2}{2f})$ . Furthermore, it can be shown that  $S_t(x)$  is both Gaussian and Markovian [9]). Being homogeneous in  $\mathcal{A}$ ,  $S_t(x)$  has the autocorrelation

$$E\{S_t(x_0)S_t(x_1)\} = \frac{\sigma^2}{2f}e^{-f(x_1-x_0)}, \quad (x_0 < x_1) \quad (2)$$

which is only a function of distance between  $x_1$  and  $x_0$ .

### 2.2. Sensor Network Model

Consider that a large number of sensors are randomly deployed in  $\mathcal{A}$ . We assume that the distribution of sensors in  $\mathcal{A}$  forms a one-dimensional homogeneous spatial Poisson field with local density  $\rho$  nodes/unit area. For an interval of size  $l$  in  $\mathcal{A}$ , the number of sensors  $N(l)$  is a Poisson random variable

$$P[N(l) = k] = e^{-\rho l} \frac{(\rho l)^k}{k!}. \quad (3)$$

Given  $N(l) = k$ , the  $k$  sensors are uniformly distributed. The number of sensors in any two disjoint intervals are independent.

The measurement of a sensor at location  $x$  and time  $t$  is given by

$$Y_t(x) = S_t(x) + N_t(x) \quad (4)$$

where  $N_t(x)$  is spatially independent and identically distributed (iid) zero mean white Gaussian measurement noise with variance  $\sigma_N^2$ . It is assumed to be independent of  $S_t(x)$ . Each sensor then stores its local measurement, along with its location information, in the form of a packet for collection. Note that we will reconstruct the source signal  $\{S_t(x) : x \in \mathcal{A}\}$  at time  $t$  based only on the sensor measurements at time  $t$ . Therefore, we drop the time index for brevity in the following presentation.

### 2.3. Two Types of MAC Schemes

We assume sensors transmit packets to the access point through a common wireless channel. We consider two MAC schemes: deterministic scheduling and random access.

A deterministic scheduling collects data from predetermined locations. One particular scheme is to schedule equally spaced sensors to transmit their packets. When there exist sensors at all

desired locations, it turns out that this is the best scheduling for the distortion measure defined below [2]. For finite sensor density, such scheme may schedule transmission from empty sensor spots. To circumvent this, we consider the scheme that enable a small interval centered at each desired location in  $\mathcal{A}$ . Details of this scheme is given in Section 3.1. We refer to this as *the deterministic scheduling*, and denote it as  $\pi_u$ .

For the random access MAC, in contrast, sensors contend to access the channel with equal priority, and their packets have equal chance to get through. Under this type of MAC, the origins of the  $M$  received packets, denoted as  $\{P_1, \dots, P_M\}$ , are random. This appears as if the access point randomly samples the sensor measurement data in  $\mathcal{A}$ . We denote this type of MAC as  $\pi_r$ .

### 2.4. Source Estimation and Reconstruction Distortion

Assume the access point receives a total of  $M$  packets (data samples) originated from some points in  $\mathcal{A}$ . To avoid the boundary effect for signal reconstruction, we assume that, during the data collection, the access point always obtains the packets from the two sensors closest to the two boundaries of  $\mathcal{A}$ . We denote the locations of this two "edge sensors" by  $P_0$  and  $P_D$ . The locations from which these  $(M+2)$  packets are received are then denoted by  $\mathbf{p}_M = \{P_0, P_1, \dots, P_M, P_D\}$ .

Given the  $M+2$  received packets (samples) from location  $\mathbf{p}_M = [P_0, P_1, \dots, P_M, P_D]$ , we denote the corresponding order statistics of the packet generation locations  $\{P_1, \dots, P_M\}$  by  $P_{(1)} < \dots < P_{(M)}$ . We estimate  $S(x)$  at  $x$  by MMSE smoothing using its two immediate neighbor samples: for  $P_{(i)} < x < P_{(i+1)}$ ,

$$\hat{S}(x) = E[S(x) | \{Y(P_{(i)}), Y(P_{(i+1)})\}]. \quad (5)$$

Let the measurement SNR be  $\text{SNR} = \sigma^2/2f\sigma_N^2$ . We define the maximum field reconstruction distortion given  $\mathbf{p}_M$  by the maximum mean square estimation error in  $\mathcal{A}$

$$\mathcal{E}(\mathbf{p}_M, \text{SNR}) \triangleq \max_{x \in \mathcal{A}} E\{|\hat{S}(x) - S(x)|^2 | \mathbf{p}_M\}. \quad (6)$$

A MAC scheme  $\pi$  specifies how packets should be transmitted, i.e., how the signal field  $\{S(x) : x \in \mathcal{A}\}$  is sampled. Therefore,  $\pi$  specifies the distribution of sample points  $\mathbf{p}_M$ . Under a given MAC scheme  $\pi$ , the average maximum distortion of the signal field is then given by

$$\bar{\mathcal{E}}(M, \text{SNR}; \pi) \triangleq E\{\mathcal{E}(\mathbf{p}_M, \text{SNR}); M, \pi\}. \quad (7)$$

where the expectation is taken over  $\mathbf{p}_M$  for a given  $M$ .

Our goal is to examine the signal field reconstruction performance under the two types of MACs. Specifically, we analyze  $\bar{\mathcal{E}}(M; \pi)$  under  $\pi_u$  and  $\pi_r$ .

## 3. RECONSTRUCTION DISTORTION: CALCULATION

### 3.1. Data Retrieval By Scheduling

In stead of scheduling transmissions from the exact equally spaced locations in  $\mathcal{A}$ , the deterministic scheduling  $\pi_u$  enables a small interval  $\epsilon$  centered at the desired location, and let one sensor in this interval to transmit. If there is no sensors in the interval, an outage occur. Let  $P_{\text{out}}$  be the outage probability of sensors. Then, the smallest possible interval  $\epsilon$ , as a function of  $P_{\text{out}}$ , should satisfy

$$P_{\text{out}} = P[N(\epsilon) = 0] = e^{-\rho\epsilon}. \quad (8)$$

i.e.,  $\epsilon = \ln 1/P_{\text{out}}/\rho$ . For Poisson distributed sensors, in none overlapping intervals, the sensor outage in each interval  $\epsilon$  are iid with probability  $P_{\text{out}}$ . The maximum number  $M_{\text{max}}$  of disjoint intervals  $\pi_u$  can enable is determined by

$$M_{\text{max}} = D/\epsilon = \frac{D\rho}{\ln 1/P_{\text{out}}}. \quad (9)$$

In summary, according to the targeted sensor outage probability  $P_{\text{out}}$  and interval length  $\epsilon$ , the scheme  $\pi_u$  divides the field into  $M$  intervals centered at  $iD/(M-1)$  for  $i = 0, \dots, M-1$ . Then, the access point collects one data sample from in each interval.

### 3.2. The Average Maximum Distortion

Suppose we want to retrieve  $M$  data samples using each MAC scheme. The average maximum distortion of  $\pi$  is then given by

$$\bar{\mathcal{E}}(M, \text{SNR}; \pi_u) = E\{\bar{\mathcal{E}}(M_o, \text{SNR}; \pi_u)\} \quad (10)$$

where the expectation is taken over  $M_o$ , the actual number of received samples.

Assume the access point receives total  $M_o$  packets. By the Gaussian property of the process  $\{S(x)\}$ , the MMSE estimator in (5) is given by

$$\hat{S}(x) = E\{S(x)\mathbf{S}_p^{(i)H}\}(E\{\mathbf{S}_p^{(i)}\mathbf{S}_p^{(i)H}\} + \sigma_n^2\mathbf{I})^{-1}\mathbf{S}_p^{(i)} \quad (11)$$

where  $\mathbf{Y}_p^{(i)} = [Y(P_{(i)}), Y(P_{(i+1)})]^T$ ,  $\mathbf{S}_p^{(i)} = [S(P_{(i)}), S(P_{(i+1)})]^T$ , and  $P_{(i)} < x < P_{(i+1)}$ . Finding  $\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR})$  in (6) can then be broken down to finding the maximum distortion of  $S(x)$  in each interval  $P_{(i)} < x < P_{(i+1)}$ , for  $i = 0, \dots, M_o$ .

For  $P_{(i)} < x < P_{(i+1)}$ , we can show that the maximum MSE of  $\hat{S}(x)$  is obtained at the mid-point of  $P_{(i)}$  and  $P_{(i+1)}$ . Moreover, it is only a function of the distance between  $P_{(i)}$  and  $P_{(i+1)}$ . Therefore, the maximum distortion  $\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR})$  in (6) is determined only by the maximum of distances between any two adjacent data samples

$$\mathcal{E}(\mathbf{p}_M, \text{SNR}) = \frac{\frac{1}{\text{SNR}} + 1 - e^{-f d_{\text{max}}^{(M_o)}}}{\frac{1}{\text{SNR}} + 1 + e^{-f d_{\text{max}}^{(M_o)}}} \frac{\sigma^2}{2f} \triangleq \mathcal{E}(d_{\text{max}}^{(M_o)}, \text{SNR}) \quad (12)$$

where  $d_{\text{max}}^{(M_o)} = \max_{0 \leq i \leq M_o+1} P_{(i)} - P_{(i-1)}$ . Then, the average maximum distortion in (7) is given by

$$\bar{\mathcal{E}}(M_o, \text{SNR}; \pi) = E\{\mathcal{E}(d_{\text{max}}^{(M_o)}, \text{SNR})\} \quad (13)$$

where the expectation is now taken over  $d_{\text{max}}^{(M_o)}$ .

For random access MAC, the locations of the  $M_o$  received packets  $\{P_1, \dots, P_{M_o}\}$  are iid random variables with uniform distribution  $U(0, D)$ . In this case, the probability distribution of the maximum sample distance  $d_{\text{max}}^{(M_o)}$ , denoted by  $F_{d_{\text{max}}}(x; M_o)$  is given by the following [2]

$$F_{d_{\text{max}}}(x; M_o) = \begin{cases} 0 & \text{if } 0 \leq x < \frac{D}{M_o+1} \\ g(M_o, x) & \text{if } \frac{D}{M_o-k+1} \leq x < \frac{D}{M_o-k}, k = 0, \dots, M_o-2 \\ 1 - (M_o+1)(1 - \frac{x}{D})^{M_o} & \text{if } \frac{D}{2} \leq x \leq D \\ 1 & \text{if } x > D. \end{cases}$$

where

$$g(M_o, x) = \sum_{i=0}^k (-1)^i \binom{M_o+1}{i} \left[ (M_o-i+1) \frac{x}{D} - 1 \right]^{M_o}.$$

Using above and (13),  $\bar{\mathcal{E}}(M, \text{SNR}; \pi_r)$  can be calculated by

$$\bar{\mathcal{E}}(M_o, \text{SNR}; \pi_r) = \int \mathcal{E}(x, \text{SNR}) dF_{d_{\text{max}}}(x; M_o). \quad (14)$$

The average maximum distortion of  $\pi_r$  in (10) is then obtained by

$$\begin{aligned} \bar{\mathcal{E}}(M, \text{SNR}; \pi_r) &= \sum_{m=0}^{M-1} \mathcal{E}(M_o, \text{SNR}; \pi_r) P(M_o = m) \\ &\quad + \mathcal{E}(M, \text{SNR}; \pi_r) P(M_o \geq m) \end{aligned} \quad (15)$$

where the first term corresponding to the case when the number of sensors in  $\mathcal{A}$  is  $M < M_o$ ; therefore, we can only collect  $M$  packets. Otherwise, as in the second term, we are always able to collect  $M_o$  packets.

For deterministic scheduling  $\pi_u$ , combining (10) and (13), the average maximum distortion  $\bar{\mathcal{E}}(M, \text{SNR}; \pi_u)$  is then given by

$$\bar{\mathcal{E}}(M, \text{SNR}; \pi_u) = E\{\mathcal{E}(d_{\text{max}}^{(M)}, \text{SNR})\} \quad (16)$$

where  $d_{\text{max}}^{(M)}$  is determined by the largest number of consecutive intervals that the sensor outage occurs. We denote this number by  $L_M$ ; it is random and depends on the sensor outage probability  $P_{\text{out}}$ . Then, (16) is bounded by

$$\begin{aligned} E[\mathcal{E}(\frac{DL_M}{M+1}, \text{SNR})] &\leq \\ \bar{\mathcal{E}}(M, \text{SNR}; \pi_u) &\leq E[\mathcal{E}(\frac{D(L_M+2)}{M_o+1}, \text{SNR})] \end{aligned} \quad (17)$$

where

$$E[\mathcal{E}(\frac{DL_M}{M+1}, \text{SNR})] = \sum_{k=0}^M \mathcal{E}(\frac{DL_M}{M+1}, \text{SNR}) P[L_M = k].$$

Calculating the distribution of  $L_M$  is essentially equivalent to finding the distribution of the longest head run (consecutive heads) in a sequence for a biased coin tossing. The distribution  $F_{L_M}(x)$  is given by

$$F_{L_M}(x) = \sum_{k=0}^M C_M^{(k)}(x) P_{\text{out}}^k (1 - P_{\text{out}})^{M-k} \quad (18)$$

where  $C_M^{(k)}(x)$  is the number of realizations in which  $k$  sensor outages occur, but no more than  $x$  of these occur consecutively. It can be shown that  $C_M^{(k)}(x)$  satisfies the following recursion

$$C_M^{(k)}(x) = \begin{cases} 0 & \text{if } k = M, x < M \\ \binom{M}{k} & \text{if } k \leq x \\ \sum_{i=0}^x C_{M-1-i}^{(k-i)}(x) & \text{if } x < k < M. \end{cases} \quad (19)$$

### 4. RECONSTRUCTION DISTORTION: PERFORMANCE COMPARISON

We now compare the reconstruction performance under  $\pi_r$  and  $\pi_u$  for different outage probability and various of  $M_{\text{max}}$  (i.e., sensor density). Fig. 2 show the probability distribution of  $L_{M_{\text{max}}}$  for

$M_{\max} = 200$  and  $P_{\text{out}} = 0.4$ . We see that  $L_{M_{\max}}$  is typically concentrated on several values. Fig. 3 plots the average maximum distortion for various  $M_{\max}$  under  $\pi_r$  and  $\pi_u$ , respectively. We set  $P_{\text{out}} = 0.4$ ,  $D = 5$ , and  $\sigma^2/2f = 1$ . We observe that as  $M_{\max}$  increases, the received data samples at the access point increases, and  $\bar{\mathcal{E}}(M_{\max}, \text{SNR}; \pi)$  under both  $\pi_r$  and  $\pi_u$  decreases. However, for the sensor outage probability we set, the random access  $\pi_r$  always results in better reconstruction performance than the deterministic scheduling  $\pi_u$  does. In Fig. 4, we plot  $\bar{\mathcal{E}}(M_{\max}, \text{SNR}; \pi)$  for various sensor outage probability  $P_{\text{out}}$ , for  $M_{\max} = 200$ ,  $D = 5$ , and  $\sigma^2/2f = 1$ . We observe that clearly there exist a threshold  $P_{th}$  on  $P_{\text{out}}$ , such that for all  $P_{\text{out}} > P_{th}$ , the performance loss under the optimal scheduling  $\pi_u$  due to missing data samples does not justify the effort of scheduling desired retrieval pattern, and simple random access outperforms optimal scheduling. Indeed, using the theory of extreme statistics, for all sufficiently large  $M_{\max}$  (i.e., high but finite sensor density), we can show that this threshold is  $P_{th} = 1/e$ .

## 5. CONCLUSION

In this paper, we consider reconstruct the Gauss Markov field under two different data retrieval strategies: the deterministic scheduling designed to collect data from equally spaced locations and the ALOHA-like random access. We have shown that, the relative reconstruction performance depends critically on the outage probability  $P_{\text{out}}$  of sensors. Our comparison demonstrates for relatively high  $P_{\text{out}}$ , the performance loss due to missing data samples does not justify the effort of scheduling desired retrieval pattern, and simple random access results in better performance.

## 6. REFERENCES

- [1] L. Tong, Q. Zhao, and S. Adireddy, "Sensor Networks with Mobile Agents," in *Proc. 2003 MILCOM*, Oct. 2003.
- [2] M. Dong, L. Tong, and B. Sadler, "Impact of MAC design on estimation of spatial Markov process in sensor networks." to be submitted to *IEEE Trans. on Signal Processing*, 2003.
- [3] W. Ye, J. Heidemann, and D. Estrin, "An Energy-Efficient MAC Protocols for Wireless Sensor Networks," in *Proc. of IEEE INFOCOM*, (New York, NY), pp. 1567–1576, June 2002.
- [4] K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Personal Communication*, pp. 16–27, October 2000.
- [5] R. Iyer and L. Kleinrock, "QoS Control for Sensor Networks," in *Proc. of 2003 ICC*, May 2003.
- [6] P. Venkitasubramaniam, S. Adireddy, and L. Tong, "Sensor Networks with Mobile Agents: Optimal Random Access and Coding," *submitted to IEEE Journal on Sel. Areas in Comm.*, July 2003.
- [7] Q. Zhao and L. Tong, "QoS Specific Information Retrieval for Densely Deployed Sensor Network," in *Proc. IEEE 2003 MILCOM*, Oct. 2003.
- [8] M. Micheli and M. I. Jordan, "Random Sampling of a Continuous-time Stochastic Dynamical System," in *Proc. of the 15th Int. Sym. on MTNS 2002*, Aug. 2002.
- [9] D. Cox and H. Miller, *The Theory of Stochastic Processes*. New York: John Wiley & Sons Inc., 1965.

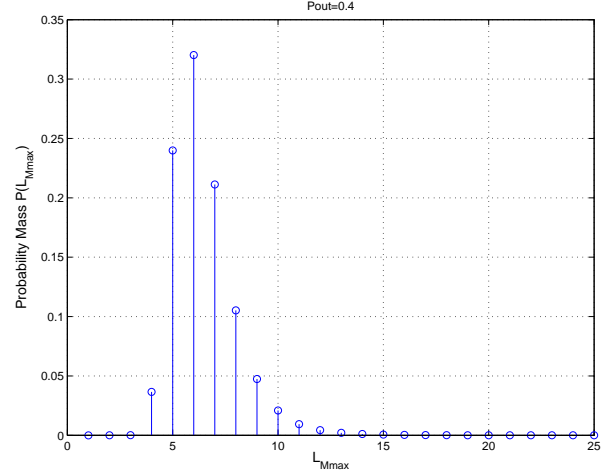


Fig. 2: PDF of  $L_{M_{\max}}$  ( $M_{\max} = 200$ )

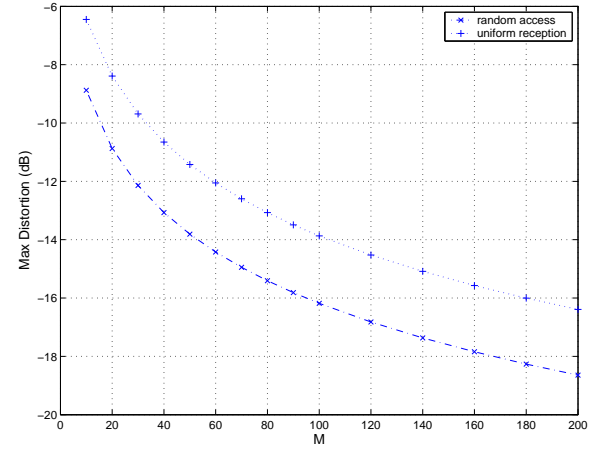


Fig. 3: Maximum distortion vs.  $M_{\max}$

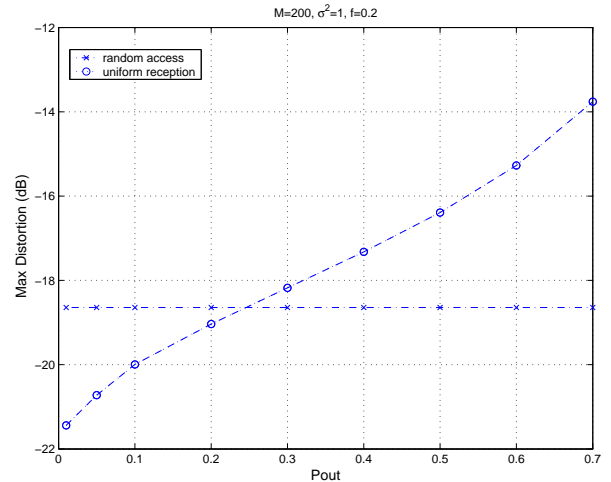


Fig. 4: Maximum distortion vs.  $P_{\text{out}}$