Distributed Adaptive Quantization and Estimation for Wireless Sensor Networks

Hongbin Li, *Member, IEEE*, and Jun Fang

***Abstract*—In this letter, the problem of distributed parameter estimation in a wireless sensor network is considered, where due to bandwidth constraint, each sensor node sends only one bit of information to a fusion center. We propose a new distributed adaptive quantization scheme by which each individual sensor node dynamically adjusts the threshold of its quantizer based on earlier transmissions from other sensor nodes. The maximum likelihood estimator (MLE) and the Cramér–Rao bound (CRB) associated with our distributed adaptive quantization scheme are derived. Numerical results depicting the performance and advantages of our approach over a fixed quantization scheme are presented.**

***Index Terms*—Adaptive quantization, distributed estimation, wireless sensor networks.**

I. Introduction

CONSIDER a wireless sensor network composed of N distributed sensor nodes. Each sensor node makes a noisy observation of an unknown parameter θ that is described by

$$x_n = \theta + w_n, \quad n = 1, 2, \ldots, N \tag{1}$$

where N denotes the number of sensors and w_n is zero mean independent and identically distributed (i.i.d.) noise with respect to (w.r.t.) n. Suppose that, due to the bandwidth/power constraint, all sensor nodes have to quantize their observations $\{x_n\}$ into one-bit binary data $\{b_n\}$. The problem of interest is to estimate θ from the quantized data $\{b_n\}$ received at the fusion center.

A number of studies have considered such a distributed estimation approach, including stochastic methods that model θ as a random parameter and require knowledge of the joint distribution of θ and the observed signals (see, e.g., [1] and [2]), as well as deterministic methods that model θ as a deterministically unknown parameter. The latter can be further classified into methods that require knowledge of the conditional distribution of x_n conditioned on θ (e.g., [3] and [4]) and methods that do not (e.g., [5] and [6]).

The achievable estimation performance at the FC can be shown to critically depend on the choice of the 1-bit quantizer that is used to quantize the data at each sensor node [4]. One strategy is to choose a *fixed* quantizer for all sensor nodes with a fixed quantization threshold τ [3]. The optimum choice of τ, however, depends on θ, which is unknown. It is found that if τ is set away from θ, the best achievable estimation performance at the FC has an exponentially increasing estimation error in $|\tau - \theta|$ [4]. An alternative strategy is to use a set of thresholds $\{\tau_k\}$, and each τ_k is used in a fraction ρ_k of the N sensor nodes [4], in the hope that some of the thresholds are close to the unknown θ. The problem is that finding the best set of $\{\tau_k, \rho_k\}$ is involved.

To deal with the above difficulty, we present a new distributed adaptive quantization scheme by which each individual sensor node dynamically adjusts the threshold of its quantizer based on earlier transmissions from other sensor nodes. Our scheme is in essence a distributed Delta modulation technique, whereby each sensor node accumulates earlier transmissions from other sensor nodes and uses the accumulated value s_{n-1} as the threshold for its 1-bit quantizer. The accumulated value s_{n-1} can be shown to converge (w.r.t. n) around the unknown θ. Based on our proposed adaptive quantization scheme, we develop the maximum likelihood estimator (MLE) that can be used at the FC to find the ML estimate of θ and the Cramér–Rao bound (CRB) that tells about the best achievable estimation performance (among all unbiased estimators) for the proposed adaptive quantizer.

II. Fixed Quantization Approach

The fixed quantization approach is to apply a common threshold τ for all sensors and generate quantized data b_n, which are sent to the FC as follows [3]:

$$b_n = \operatorname{sgn}(x_n - \tau), \quad n = 1, 2, \ldots, N. \tag{2}$$

The probability mass function (PMF) of b_n is given by

$$P(b_n; \theta) = [F_w(\tau - \theta)]^{(1+b_n)/2}[1 - F_w(\tau - \theta)]^{(1-b_n)/2} \tag{3}$$

where $F_w(x)$ denotes the complementary cumulative density function (CCDF) of w_n. Since $\{b_n\}$ are i.i.d., the log-PMF or log-likelihood function is shown in (4) at the bottom of the next page, where the subscript FQ is used to denote the fixed quantization scheme. The MLE is given by [4]

$$\begin{aligned}\hat{\theta}_{\mathrm{FQ}} &= \arg\max_{\theta} L_{\mathrm{FQ}}(\theta) \\ &= \tau - F^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\frac{1+b_n}{2}\right).\end{aligned} \tag{5}$$

The CRB based on the above fixed quantization is [3], [4]

$$\mathrm{CRB}_{\mathrm{FQ}}(\theta) = \frac{F_w(\tau - \theta)[1 - F_w(\tau - \theta)]}{N p_w^2(\tau - \theta)} \tag{6}$$

where $p_w(x)$ denotes the probability density function (PDF) of w_n. We see that $\mathrm{CRB}_{\mathrm{FQ}}(\theta)$ depends on the threshold τ. Fur-

Manuscript received January 29, 2007; revised February 27, 2007. This work was supported in part by the National Science Foundation under Grant CCF-0514938 and in part by the U.S. Army ARDEC. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Richard J. Kozick.

The authors are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu; Jun.Fang@stevens.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2007.896390

thermore, it has been found that the CRB increases exponentially with $|\tau - \theta|/\sigma$ for the Gaussian noise, where σ denotes the standard deviation of the noise [4].

III. Distributed Adaptive Quantization Approach

In this section, we first introduce our distributed adaptive quantization scheme, followed by the development of the MLE and CRB.

A. Adaptive Quantization

We assume that the sensors share the communication channel on a time-sharing basis (e.g., each sensor is polled by the FC), so that sensor 1 transmits first, followed by sensor 2, and so on and so forth. The 1-bit quantizer at sensor 1 uses a zero-threshold to generate b_1 as follows:

$$b_1 = \text{sgn}\{x_1\}. \tag{7}$$

Then, b_1 is sent (i.e., broadcast) to the FC as well as the other $N-1$ sensors. After receiving b_1, sensor 2 computes $s_1 = \Delta b_1$, where Δ is a step size parameter whose choice is discussed later, and generates b_2 as follows:

$$b_2 = \text{sgn}\{x_2 - s_1\}. \tag{8}$$

In general, for sensor n, it first forms a cumulative sum

$$s_{n-1} = s_{n-2} + \Delta b_{n-1} = \Delta \sum_{k=1}^{n-1} b_k \tag{9}$$

and then, it uses s_{n-1} as a threshold for quantization

$$b_n = \text{sgn}\{x_n - s_{n-1}\}. \tag{10}$$

One can immediately recognize that the above process is reminiscent of the Delta modulation but is implemented in a distributed fashion.

B. MLE

Different from the fixed quantization approach, the binary data bits $b_1, b_2, \ldots, b_N$ generated by our distributed adaptive quantization are no longer independent and identically distributed. The conditional PMF of b_n is given by

$$\begin{aligned} P(b_n|b_1,\ldots,b_{n-1};\theta) = {} & [F_w(s_{n-1}-\theta)]^{(1+b_n)/2} \\ & \times [1 - F_w(s_{n-1}-\theta)]^{(1-b_n)/2}. \end{aligned} \tag{11}$$

Using conditional probabilities, we can write the joint PMF of $b_1, b_2, \ldots, b_N$ as

$$\begin{aligned} P(b_1,\ldots,b_N;\theta) &= \prod_{n=1}^{N} P(b_n|b_1,\ldots,b_{n-1};\theta) \\ &= \prod_{n=1}^{N} P(b_n|s_{n-1};\theta). \end{aligned} \tag{12}$$

Then the log-likelihood function following from (11) and (12) is shown in (13) at the bottom of the page, where the subscript AQ is used to denote our adaptive quantization scheme. As such, the MLE is given by

$$\hat{\theta}_{\text{AQ}} = \arg\max_{\theta} L_{\text{AQ}}(\theta). \tag{14}$$

Unlike $\hat{\theta}_{\text{FQ}}$, an analytical form of $\hat{\theta}_{\text{AQ}}$ cannot be obtained. Nevertheless, we can employ the gradient-based iterative algorithm to find $\hat{\theta}_{\text{AQ}}$. When the noise is Gaussian distributed, it can be shown that the MLE function in (14) is concave and guaranteed to converge to the global maximum (see, e.g., [4, Proposition 2]).

C. CRB

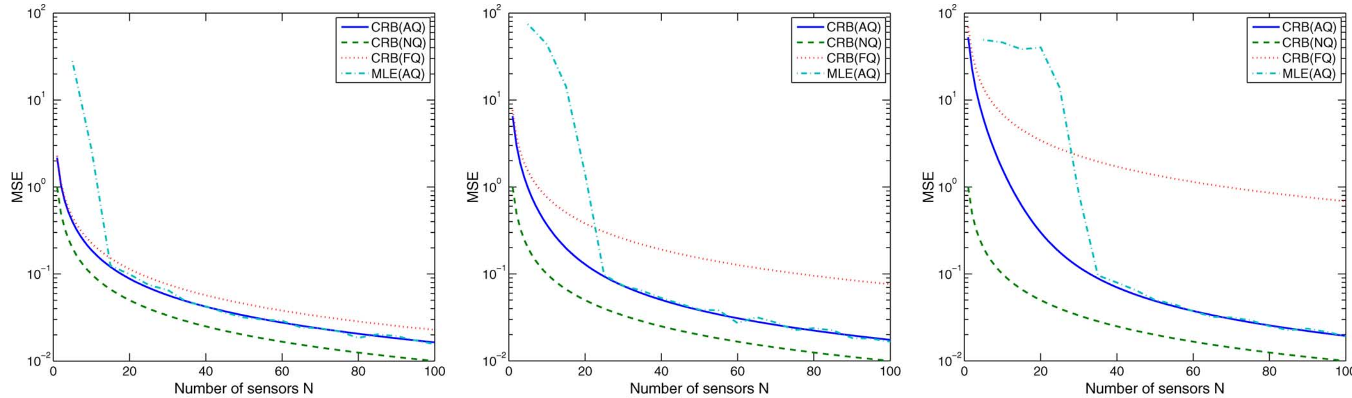
Noting that $F'_w(x) \triangleq \partial F_w(x)/\partial x = -p_w(x)$, we can quickly verify that the second-order derivative of $L_{\text{AQ}}(\theta)$ is shown in (15) at the bottom of the next page, where $p'_w(x) \triangleq \partial p_w(x)/\partial x$. The Fisher information for the estimation problem is given by (e.g., [7])

$$\begin{aligned} J_{\text{AQ}}(\theta) &= -E\left\{\frac{\partial^2 L_{\text{AQ}}(\theta)}{\partial\theta^2}\right\} \\ &= -\sum_{n=1}^{N} E_{s_{n-1}}\left\{E_{b_n|s_{n-1}}[A(b_n, s_{n-1}, \theta)]\right\} \\ &\overset{(a)}{=} \sum_{n=1}^{N} E_{s_{n-1}}\left[\frac{p_w^2(s_{n-1}-\theta)}{F_w(s_{n-1}-\theta)(1-F_w(s_{n-1}-\theta))}\right] \end{aligned} \tag{16}$$

where $E_{s_{n-1}}$ denotes the expectation w.r.t. the distribution $P(s_{n-1};\theta)$, $E_{b_n|s_{n-1}}$ denotes the expectation w.r.t. the conditional distribution $P(b_n|s_{n-1};\theta)$, and (a) follows from the fact that b_n is a binary random variable with $P(b_n = 1|s_{n-1},\theta) = F_w(s_{n-1}-\theta)$ and $P(b_n = -1|s_{n-1},\theta) = 1 - F_w(s_{n-1}-\theta)$. Note that s_{n-1} is a discrete random variable with possible values

$$\begin{aligned} L_{\text{FQ}}(\theta) &\triangleq \ln[P(b_1,\ldots,b_N;\theta)] \\ &= \sum_{n=1}^{N}\left\{\left(\frac{1+b_n}{2}\right)\ln[F_w(\tau-\theta)] + \left(\frac{1-b_n}{2}\right)\ln[1-F_w(\tau-\theta)]\right\} \end{aligned} \tag{4}$$

$$L_{\text{AQ}}(\theta) = \sum_{n=1}^{N}\left\{\left(\frac{1+b_n}{2}\right)\ln[F_w(s_{n-1}-\theta)] + \left(\frac{1-b_n}{2}\right)\ln[1-F_w(s_{n-1}-\theta)]\right\} \tag{13}$$



(a) (b) (c)

Fig. 1. CRB and MLE versus the number of sensors N. (a) $\theta = 1$. (b) $\theta = 2$. (c) $\theta = 3$.

$\{\tau_{-N}, \ldots, \tau_{-1}, 0, \tau_1, \ldots, \tau_N\}$ for any $n \le N$, where $\tau_k \triangleq k\Delta$. Hence, we can further express (16) as

$$\begin{aligned} J_{\mathrm{AQ}}(\theta) &= \sum_{n=1}^{N}\sum_{k=-N}^{N} P(s_{n-1}=\tau_k)\left[\frac{p_w^2(\tau_k-\theta)}{F_w(\tau_k-\theta)(1-F_w(\tau_k-\theta))}\right] \\ &= \sum_{k=-N}^{N}\sum_{n=1}^{N} P(s_{n-1}=\tau_k)\left[\frac{p_w^2(\tau_k-\theta)}{F_w(\tau_k-\theta)(1-F_w(\tau_k-\theta))}\right] \\ &= N\sum_{k=-N}^{N}\left[\frac{\rho_k p_w^2(\tau_k-\theta)}{F_w(\tau_k-\theta)(1-F_w(\tau_k-\theta))}\right] \end{aligned} \quad (17)$$

where $\rho_k \triangleq (1/N)\sum_{n=1}^{N} P(s_{n-1} = \tau_k)$, which denotes the normalized frequency with which the sensors use τ_k as the threshold. Therefore, the CRB is given by

$$\begin{aligned} \mathrm{CRB}_{\mathrm{AQ}}(\theta) &= \frac{1}{J_{\mathrm{AQ}}(\theta)} \\ &= \frac{1}{N}\left[\sum_{k=-N}^{N}\frac{\rho_k p_w^2(\tau_k-\theta)}{F_w(\tau_k-\theta)(1-F_w(\tau_k-\theta))}\right]^{-1}. \end{aligned} \quad (18)$$

Interestingly, (18) has the same analytical form as that in [4] for the nonidentical thresholds case (see [4, (17)]). Nevertheless, the difference between our scheme and that of [4] is obvious. First, the frequencies $\{\rho_k\}$ associated with the thresholds $\{\tau_k\}$ are computed in advance in [4] while these parameters are induced in an adaptive way in our work. Second, in [4], a fixed threshold is assigned to a single sensor; however, for our scheme, each sensor has a random threshold with a certain distribution.

For w_n with a symmetric PDF, if Δ is chosen small enough, the random walk process $\{s_{n-1}\}$ converges around the unknown parameter θ as n becomes large. This is because

$$\begin{aligned} E_{s_n|s_{n-1}}[s_n] &= (s_{n-1}+\Delta)F_w(s_{n-1}-\theta) \\ &\quad + (s_{n-1}-\Delta)(1-F_w(s_{n-1}-\theta)) \\ &= \begin{cases} > s_{n-1}, & \text{if } s_{n-1} < \theta \\ = s_{n-1}, & \text{if } s_{n-1} = \theta \\ < s_{n-1}, & \text{if } s_{n-1} > \theta \end{cases} \end{aligned} \quad (19)$$

which indicates that $E[s_n]$ tends toward θ with increasing n. For Gaussian noise, it is clear that (18) is minimized if only one threshold $\tau = \theta$ is employed, i.e., $\tau_k = \theta$ for any k. While this is impossible (since θ is unknown), our approach provides a more practical solution: it automatically selects the thresholds $\{\tau_k\}$ and the corresponding frequencies $\{\rho_k\}$ such that for sufficiently large N and proper step size Δ, the thresholds $\{\tau_k\}$ are around the unknown θ with a high probability. This is notably different from [4], where $\{\rho_k\}$ is computed using a weighting function that is in general difficult to choose.

To compute the CRB in (18), we need the distribution of s_{n-1}. Note that s_{n-1} is a random walk process with varying probabilities of the increment of Δ and $-\Delta$. Although finding a closed-form expression for the PMF of s_{n-1} seems cumbersome, it can be computed rather straightforwardly by recursive calculation. For simplicity, let $P_{i,j} \triangleq P(s_i = j\Delta)$. Then the distribution of s_{n-1} can be calculated recursively as

$$\begin{aligned} P_{i,j} &= P_{i-1,j-1}P(b_i=1) + P_{i-1,j+1}P(b_i=-1) \\ &= P_{i-1,j-1}F_w((j-1)\Delta-\theta) \\ &\quad + P_{i-1,j+1}[1-F_w((j+1)\Delta-\theta)]. \end{aligned} \quad (20)$$

It is clear that $\mathrm{CRB}_{\mathrm{AQ}}$ depends on the step size parameter Δ because the choice of Δ affects $\{\tau_k\}$ and $\{\rho_k\}$ [see (18)].

$$\begin{aligned} \frac{\partial^2 L_{\mathrm{AQ}}(\theta)}{\partial\theta^2} &= \sum_{n=1}^{N}\left\{\left(\frac{1+b_n}{2}\right)\left(-\frac{p_w'(s_{n-1}-\theta)}{F_w(s_{n-1}-\theta)} - \frac{p_w^2(s_{n-1}-\theta)}{F_w^2(s_{n-1}-\theta)}\right) - \left(\frac{1-b_n}{2}\right)\left(-\frac{p_w'(s_{n-1}-\theta)}{\left[1-F_w(s_{n-1}-\theta)\right]} + \frac{p_w^2(s_{n-1}-\theta)}{\left[1-F_w(s_{n-1}-\theta)\right]^2}\right)\right\} \\ &\triangleq \sum_{n=1}^{N} A(b_n, s_{n-1}, \theta) \end{aligned} \quad (15)$$

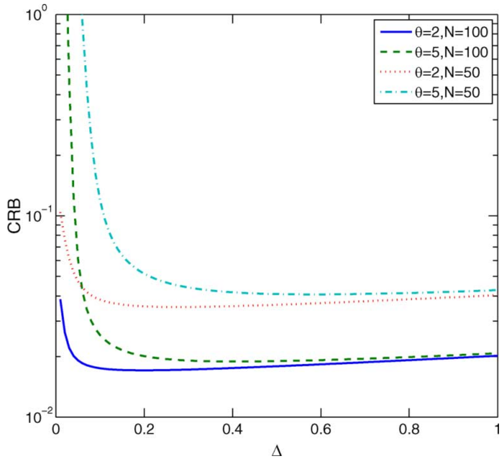


Fig. 2. CRB versus Δ with $\theta = 2, 5$ and $N = 50, 100$, respectively.

Defining $h(x) \triangleq p_w^2(x)/(F_w(x)(1 - F_w(x)))$, we express J_{AQ} (the inverse of $\mathrm{CRB}_{\mathrm{AQ}}$) as

$$J_{\mathrm{AQ}} = N \sum_{k=-N}^{N} \rho_k h(x_k) \tag{21}$$

where $x_k \triangleq \tau_k - \theta$. The choice of Δ is discussed next. Our discussion applies to any distribution that leads to a unimodal $h(x) \triangleq p_w^2(x)/F_w(x)(1 - F_w(x))$ achieving its maximum at $x = 0$. Examples of such distributions include the Gaussian, Laplacian, and Cauchy PDF, etc. In order to maximize J_{AQ} (minimize $\mathrm{CRB}_{\mathrm{AQ}}$), Δ should be chosen to make the coefficients $\{\rho_k\}$ as large as possible for those k whose corresponding values $\{\tau_k\}$ are close to θ. If Δ is chosen too small such that $\Delta \leq |\theta|/N$, the coefficients $\{\rho_k\}$ have its major values scattered while k varies from $-N$ to N and cannot form a peak around k_0, where $x_{k_0} = 0$. On the other hand, if Δ is chosen too large, although the coefficients $\{\rho_k\}$ have peak values around k_0, $\{h(x_k)\}$ will become small as a large Δ causes a large deviation of x_k from 0. Thus, the choice of Δ should be made with a tradeoff between these two opposite effects.

IV. Numerical Results and Discussions

To illustrate the performance of the proposed distributed adaptive quantization and estimation scheme, we consider the case where the sensor noise $\{w_n\}$ are independent and identically distributed Gaussian random variables with zero mean and variance $\sigma^2 = 1$. We compare our adaptive scheme with the fixed quantization approach described in Section II and the clairvoyant approach that uses unquantized data. The CRB for the clairvoyant approach provides a benchmark (lower bound) for comparison and is given by [4]

$$\mathrm{CRB}_{\mathrm{NQ}}(\theta) = \frac{\sigma^2}{N} \tag{22}$$

where the subscript NQ denotes that no quantization is used.

Fig. 1(a)–(c) shows the CRB of the above three approaches when $\theta = 1$, 2, and 3, respectively. For the fixed quantization approach, we set the threshold $\tau = 0$, and for our adaptive quantization approach, we choose $\Delta = 0.1$. As we can see, the fixed quantization approach is sensitive to the value of θ or, equivalently, the value of τ; as the two become more apart from each other (even not too far apart), the performance of the fixed quantization approach degrades significantly. On the other hand, our adaptive quantization scheme does not have the above problem. In addition, in all three cases considered, it outperforms the fixed quantization scheme and is closer to the clairvoyant approach. The mean-square error (MSE) of the adaptive quantization-based maximum likelihood estimator (MLE) is also included in Fig. 1(a)–(c). We observe that the MSE approaches the corresponding CRB for large N.

Fig. 2 shows the CRB of the adaptive quantization scheme under different choices of Δ. It is easy to see that the optimal choice of the parameter Δ is related to the unknown parameter θ and the number of sensors N. This can be intuitively justified because a larger θ (given a fixed N) requires a larger step size to come close to the unknown parameter; likewise, a smaller N (given a fixed θ) requires a larger Δ to achieve the same effect. Another observation made on the figure is that, centered on the optimal point of Δ, the performance has a sharp degradation with a decreasing Δ while it varies very slowly with an increasing Δ. This suggests us to use an overestimated instead of an underestimated Δ.

V. Conclusion

We have proposed an AQ-based distributed estimation scheme that is robust to the unknown parameter and outperforms the fixed quantization approach. Our scheme, as well as those in [3]–[6], transmit only 1 data bit per sensor. A natural question is why to send a single bit given that tens or more overhead bits are usually required to set up a communication link. Nevertheless, our results show that for sufficiently large N, the gap between our scheme and the clairvoyant NQ is fairly small; unless higher accuracy is desirable, there is no need to transmit more bits. It should be noted that our method can be readily extended to incorporate multi-bit quantization as well: each sensor just uses prior transmissions from other sensors to adjust the center threshold of its multi-bit quantizer. Future research work may include adaptive adjustment of the step size Δ, which is expected to yield additional performance gain, and having each sensor to broadcast to its neighbors only, in order to reduce communication overhead and energy cost.

References

[1] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 7, pp. 1456–1459, Jul. 1993.
[2] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 181–184, Feb. 2000.
[3] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.
[4] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, Part I Gaussian PDF," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.
[5] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.
[6] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.
[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.