# SADDLEPOINT APPROXIMATION FOR SENSOR NETWORK OPTIMIZATION

*Saeed A. Aldosari, José M. F. Moura*

Department of Electrical and Computer Engineering, Carnegie Mellon University
Pittsburgh, U.S.A
{saldosar, moura}@ece.cmu.edu

## ABSTRACT

The task of detection optimization in sensor networks is hindered by the large computational cost of evaluating the performance criteria, e.g. the probability of making wrong decisions. We present an approach that avoids these obstacles by considering a rather accurate approximation to computing the detection performance. We propose the saddlepoint approximation and provide results that demonstrate its high accuracy and low complexity. The results are used to show that, for a range of problems, the optimal fusion rule is equivalent to a simple majority rule.

## 1. INTRODUCTION

The design of reliable and efficient sensor networks is attracting attention in recent years. Many factors are important in designing the sensor network, including the sensing performance, network lifetime, or communication bandwidth. Coupling this with the expanding network size leads usually to ill-conditioned optimization problems requiring extremely large computational resources. In this paper, we consider the problem of *globally* optimizing sensor networks with respect to the error detection performance. We consider a parallel-fusion architecture in which all sensor nodes send their local decisions to a fusion center. The aim is to develop a practical approach for optimizing both the local detectors at the sensor nodes and the (global) fusion rule such that the overall detection error is minimized.

Finding the global fusion rule is essentially a combinatorial problem that requires repeated evaluation of the performance measure. For a given fixed network structure, i.e., given set of local detectors and fixed global fusion rule, the computation of the performance criterion should be both, fast and accurate. Unfortunately, the average probability of error criterion that is usually adopted is extremely costly to compute. In this paper, we develop an approximation to the average probability of error that is both simple to compute and accurate. We adopt the *saddlepoint* approximation [1], which has been used in many applications such as optical detection, bootstrapping, and queuing analysis. We use the Lugannani-Rice version, which is one of the most popular forms of the saddlepoint approximation [1]. In the sequel, we will refer to it interchangeably as the Lugannani-Rice or the saddlepoint approximation.

Although based on asymptotic expansions, the saddlepoint approximation is highly accurate even for small problems (e.g., networks with few number of sensors). In addition, the saddlepoint approximation avoids complexity bottlenecks since it is based on
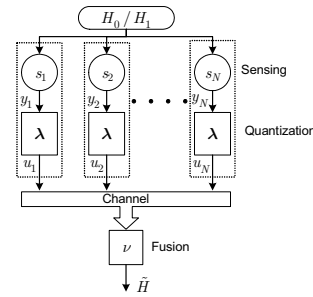
**Fig. 1**: Parallel fusion network.

only a few simple expressions. This contrasts with exact evaluation methods, which have complexity of order $N^{2^b-1}$, where $N$ is the number of sensors and $b$ is the number of bits quantizing the local decisions at each sensor. The simplicity of the saddlepoint approximation enables us to address the problem of network design, in particular, of finding the optimal global fusion rule. Due to space limitations, we will only show in this paper some results on network optimization, presenting the details elsewhere.

We study the accuracy of the saddlepoint approximation by showing the impact of the conditions under which the decision variables become lattice variables. The saddlepoint approximation is then used to globally optimize a network of 100 sensors. We show that, when the noise is Laplacian (a heavy tail distribution), the optimal fusion rule reduces to a very simple and intuitive rule—the majority rule, that generalizes to the $M$-ary problem the usual binary majority rule. This may not be the case for other types of noise, for example, for Gaussian noise, except for small networks such as those considered in [2].

In section 1, we present the model and state the problem. In section 2, we present the optimal fusion rule and discuss the statistical distribution of the decision variables. Ways for evaluating the exact detection performance along with their complexity are presented in section 3. Section 4 describes the saddlepoint approximation and presents bounds for its accuracy. Finally, we present numerical results and concluding remarks in section 6.

## 2. MODEL AND PROBLEM STATEMENT

Fig. 1 depicts a network of $N$ sensors collecting $N$ scalar measurements $y_1, y_2, \ldots, y_N$ in an environment that takes one of two possible states, $H_0$ and $H_1$. Each sensor processes its own measurement independently using a $T$-threshold scalar quantizer to form a local $M$-ary decision $u_n \in \{1, 2, \ldots, M\}$, $n = 1, 2, \ldots, N$.

Each local classification is encoded into $b = \lceil \log_2 M \rceil$ bits then the local messages from all sensors are transmitted to a single fusion center. The fusion center makes the final binary decision $\widetilde{H}$ about the true state $H \in \{0, 1\}$. Our aim in this binary detection problem is to optimize the local quantizers and the fusion rule such that the average probability of error $P_e = \Pr(\widetilde{H} \neq H)$ is minimized.

We consider a Bayesian formulation in which the probabilities $\pi_0 = \Pr(H_0)$ and $\pi_1 = \Pr(H_1)$ are known *apriori*. The local measurements $y_n$ are assumed to be independent and identically distributed (i.i.d) with known density functions $f_0(y) = f(y|H_0)$ and $f_1(y) = f(y|H_1)$ under $H_0$ and $H_1$, respectively. We further assume that the likelihood ratio $f_1(y)/f_0(y)$ is monotonic in $y$. Under these assumptions, optimal local decision rules are nothing but scalar quantizers of the real measurements $y_n$.

We also assume that the local quantizers are identical. While this might not be necessarily optimal, Tsitsiklis in [3] showed that the performance loss due to this assumption is negligible when the number of sensors is large. Also, the numerical study in [2] shows that this statement holds even for small networks. We assume that the number of quantization thresholds $T = M - 1$. In this case, the "identical quantizers" assumption implies that all sensor nodes share the same vector of quantization thresholds $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_T)$, $\lambda_1 < \lambda_2 < \cdots < \lambda_T$. Extension to the general case $(T > M - 1)$ is straightforward, though at the expense of computational complexity.

### 3. OPTIMAL FUSION

Given the above assumptions, the optimal fusion rule for a particular local classification vector $\mathbf{u} = (u_1, u_2, \ldots, u_N)$ is essentially a likelihood ratio test of the form [4]

$$S_N = \sum_{n=1}^{N} \ell_n \underset{0}{\overset{1}{\gtrless}} \log \frac{\pi_0}{\pi_1} \triangleq \nu, \tag{1}$$

where $\ell_n = \log \frac{\Pr(u_n|H_1)}{\Pr(u_n|H_0)}$ is the log likelihood ratio (LLR) of making a decision $u_n$, and $\nu$ is the fusion threshold. The LLR $\ell_n$ is a discrete random variable that takes one of $M$ possible values $L_m$, $m = 1, 2, \ldots, M$, and has a conditional probability mass function (pmf) given by

$$P_{mi} = \Pr(\ell_n = L_m | H_i) = \int_{\lambda_{m-1}}^{\lambda_m} f_i(y) dy$$

under hypothesis $H_i$, where $\lambda_0 = -\infty$, $\lambda_M = +\infty$, and $L_m = \log(P_{m1}/P_{m0})$ is the LLR of an $m$ classification (making a local decision in favor of $m$).

The decision rule in (1) can also be written as a weighted majority rule

$$\sum_{m=1}^{M} N_m L_m \underset{0}{\overset{1}{\gtrless}} \nu, \tag{2}$$

where $N_m$ is the number of sensors making a $u_n = m$ classification and $L_m$ is the LLR value of that classification as defined above. Note that, for predefined local quantization thresholds, the LLRs $L_m$, $m = 1, 2, \ldots, M$, can be computed from the measurement densities. On the other hand, $N_m \in \{0, 1, \ldots, N\}$, $m = 1, 2, \ldots, M$, are discrete random variables with unknown distributions, but their sum should be $N$.

### 4. EXACT PROBABILITY OF ERROR

A globally optimal detector minimizes the average probability of error $P_e = \pi_0 P_{e0} + \pi_1 P_{e1}$ where $P_{ei} = \Pr(\widetilde{H} \neq H|H_i)$ is the probability of error under hypothesis $H_i$. One approach for computing $P_{ei}$ uses the fact that the fusion center only cares about the counts of distinct decisions as in (2). The probability of error under $H_i$ is given by

$$P_{ei} = \sum_{N_1, N_2, \ldots N_M} \left[ \prod_{m=1}^{M} (P_{mi})^{N_m} \right] \left[ \prod_{m=1}^{M} \binom{N - \sum_{k=1}^{m-1} N_k}{N_m} \right]$$

$$\text{such that } \sum_{m=1}^{M} N_m = N \text{ and } \sum_{m=1}^{M} N_m L_m \underset{\text{if } i=1}{\overset{\text{if } i=0}{\gtrless}} \nu. \tag{3}$$

The sum in the last equation considers all possible ways of selecting $M$ integers such that their sum is equal to $N$. The number of such combinations is equal to $\binom{N+M-1}{N}$. Using Stirling's approximation and assuming $N \gg M$, the number of terms in (3) is on the order of $N^{M-1}$. This is not appropriate for practical use except when $M$ and $N$ are small. The difficulties become even more prominent when we consider using such exact formulas in sensor network optimization.

### 5. SADDLEPOINT APPROXIMATION

Saddlepoint techniques are powerful tools to derive accurate approximations to densities and tail probabilities of sums of random variables. In this section, we use the saddlepoint technique to approximate the tail probabilities of $S_N = \sum_{n=1}^{N} \ell_n$ in order to compute the average probability of error $P_e$. In what follows, we drop the subscript from $\ell_n$ and use $\ell$ instead. Before presenting the approximation we need to define the cumulant generating function and its derivatives, which are used in the saddlepoint approximation. Under hypothesis $H_i$, we define the moment generating function of $\ell$ given by

$$G(\theta) = E\left(e^{\theta \ell}\right) = \sum_{m=1}^{M} P_{mi} e^{\theta L_m},$$

where $E$ denotes the expectation. The cumulant generating function of $\ell$ and its first and second derivatives are given by

$$K(\theta) = \log G(\theta),$$
$$K'(\theta) = \partial K(\theta)/\partial \theta = W_1(\theta)/G(\theta),$$
$$K''(\theta) = \partial^2 K(\theta)/\partial \theta^2 = \left[G(\theta)W_2(\theta) - W_1(\theta)^2\right]/G(\theta),$$
$$W_k(\theta) = \sum_{m=1}^{M} P_{mi}(L_m)^k e^{\theta L_m}.$$

The Lugannani-Rice formula is one of the most popular and easy forms of the saddlepoint approximation. In the context of our problem, the Lugannani-Rice formula approximates the error performances $P_{e0}$ and $P_{e1}$ as follows

$$P_{e0} = \Pr(S_n \geq \nu) \simeq \Phi(r) + \varphi(r)\left[q^{-1} - r^{-1}\right],$$
$$P_{e1} = \Pr(S_n \leq \nu) \simeq \Phi(r) - \varphi(r)\left[q^{-1} - r^{-1}\right], \tag{4}$$

$$r = \text{Sgn}(\hat{\theta})\sqrt{2(\hat{\theta}\nu - NK(\hat{\theta}))}, \tag{5}$$
$$q = \hat{\theta}\sqrt{NK''(\hat{\theta})}, \tag{6}$$

where $\varphi(x)$ is the normal density function, $\Phi(x) = \int_x^\infty \varphi(y)dy$, is the normal right-tail distribution, $\mathrm{Sgn}(x)$ denotes the sign of $x$ and $\hat\theta$ is the saddlepoint given by the unique solution to the saddlepoint equation $K'(\hat\theta) = \nu/N$, which can be solved using standard numerical univariate root finding algorithms. Care should be taken when evaluating the various quantities in the above expressions as it should be carried out under the appropriate hypothesis. In other words, the values of $r$, $q$, $K(\hat\theta)$, $K''(\hat\theta)$, and the saddlepoint $\hat\theta$ used to evaluate $P_{e0}$ are different from those of $P_{e1}$.

The form presented in (4), (5), and (6) is often used to approximate the tail probabilities of sums of continuous random variables. For this reason we refer to it as LR-Cont (i.e., the continuous form of the Lugannani-Rice approximation). However, the problem that we are considering involves the sum of discrete random variables $\ell_n$. So, the question is: is it still a good approximation to $P_e$ when the random variables are discrete? Applying the same approximation above for discrete random variables has been justified by Booth *et al.* [5] by showing that the relative error of the approximation decays rapidly as the number of samples $N$ grows. It is shown in [5] that, in almost all cases, the *relative* error is $o(N^{-1/2})$ or $O(N^{-1})$ when $M = 5, 6$ or $M > 6$, respectively ($f = o(g)$ means that $f/g \to 0$ when $N \to \infty$, while $f = O(g)$ means that $f/g$ is bounded when $N \to \infty$). Recall that the sensor nodes produce local decisions based on local $M$-ary classifiers where $M$ is the number of local quantization levels given by $M = 2^b$ and $b$ is the number of bits per sensor. No similar justification has been shown in [5] for the cases of $M = 3, 4$ (corresponding to ternary and quaternary sensor nodes, respectively) although numerical results in the following section show that the approximation performs well for those cases also. The case of $M = 2$ (binary sensor nodes) is a little different since the distribution of $\ell$ becomes always lattice valued as we explain next.

The log likelihood ratio random variable $\ell$ is lattice distributed when every possible value of it ($L_m$, $m = 1, \ldots, M$) can be represented in the form $\alpha + \beta z$, where $z$ is an integer and $\alpha, \beta \neq 0$. When $M = 2$ (binary sensors), it is easy to see that $\ell$ is always lattice regardless of the values of $L_1$ and $L_2$. When $M > 2$, the distribution of $\ell$ can also be lattice, but only under specific conditions on the noise distribution and the quantization thresholds. The reason for raising the issue of lattice versus non-lattice conditions is two fold. First, it can be shown that when $\ell$ is lattice the fusion rule in (1) can be replaced with a simple majority rule, which makes its decisions based on the integer sum of the received local classifications $u_n \in \{1, 2, \ldots, M\}$ as follows

$$\sum_{n=1}^N u_n \underset{0}{\overset{1}{\gtrless}} \log \frac{\pi_1}{\pi_0} - \frac{N}{\beta}(L_1 + \beta) \triangleq \kappa. \quad (7)$$

This brings another complexity reduction, which may be necessary for power and complexity constrained sensors nodes. Second, when the random variable $\ell$ is lattice-valued there are other forms of the saddlepoint approximation that are specific for lattice-valued variables. One such approximation is obtained by using the same equations as before ((4) and (5)) but where $q$ is replaced now with

$$q = \frac{1}{\beta}(1 - e^{-\beta\hat\theta})\sqrt{NK''(\hat\theta)}, \quad (8)$$

where $\beta$ is the lattice span. This particular form has a relative error of $O(n^{-1})$ for any $M$ provided that $\ell$ is lattice distributed. We refer to this approximation as LR-Latt (i.e., the lattice form of the Lugannani-Rice approximation). This approximation is valid only at the lattice edges. For example, in the following section, we consider a lattice case in which $M = 4$ (quaternary sensors) and
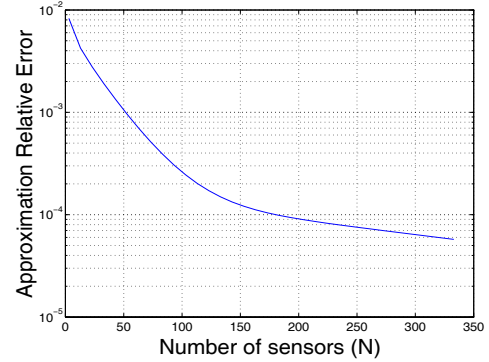


**Fig. 2**: Relative error of the saddlepoint approximation as a function of the number of sensors.

the values of $L_m$, $m = 1, \ldots, 4$, are symmetric about zero with a span of $\beta$. To approximate $P_{e0} = \Pr(S_n \geq 0)$ for that example, the LR-Latt formula should be evaluated at $\nu = \beta/2$.

In addition to its high accuracy, the saddlepoint approximation is much more practical than the exact approach. Evaluating the probability of error given the local quantization threshold $\boldsymbol{\lambda}$ requires computing few simple expressions in addition to finding the saddlepoint $\hat\theta$, which can be obtained using numerical univariate techniques, a much simpler task than the exact evaluation in Section 4, which has a computational complexity on the order of $N^{M-1}$, where $M = 2^b$. We use this fast, accurate, and simple saddlepoint approximation in an optimization algorithm that searches for the optimal local quantization thresholds $\boldsymbol{\lambda}$ as will be illustrated in the next section. Recall that, for the problem considered in this paper, optimality of the fusion rule is ensured by using the likelihood-ratio test in (2).

## 6. NUMERICAL RESULTS

We present numerical results to illustrate the accuracy of the proposed approximation and clarify the difference between the lattice and the non-lattice conditions. We also present an application to the problem of global network optimization. To model the observations, we consider an additive noise model $y = \mu_i + \vartheta$, where $\mu_i$ is the signal mean under $H_i$, $i = 0, 1$ and $\vartheta$ is a zero-mean noise with known distribution and variance $\sigma^2$. Here, we assume $\mu_1 = -\mu_0 = \mu = 1$ and that the hypotheses are equally likely to occur, i.e., $\pi_0 = \pi_1 = 1/2$. A network of $N$ quaternary ($M = 4$) sensors is used to detect the true state of nature.

**Example 1** We illustrate the accuracy of the proposed approximation in a low-SNR environment. The noise is assumed to be Gaussian and the SNR is $-10$ dB (SNR$= 10\log_{10} \mu^2/\sigma^2$). The local quantization thresholds are fixed at $\boldsymbol{\lambda} = (-3, 0, 3)$. Fig. 2 shows how the relative error of the LR-Cont approximation decays as the size of the network grows and that it performs well even for very small networks (relative error is less than 1%).

**Example 2** Here, we demonstrate the difference between lattice and non-lattice conditions and evaluate the performance of both forms of the saddlepoint approximation. The noise is assumed to be Laplacian with variance $\sigma^2 = 10$ (SNR$= -10$ dB). The network size is fixed at $N = 101$ while the local quantization thresholds are varied to examine situations where the distribution falls into a lattice. Specifically, a symmetric threshold vector $\boldsymbol{\lambda} = (-\lambda, 0, \lambda)$ is used and the value of $\lambda$ is varied from 0 to 1. In implementing the LR-Latt approximation, we assume the span of the lattice to be $\beta = L_2 - L_1$. Fig. 3 compares the relative
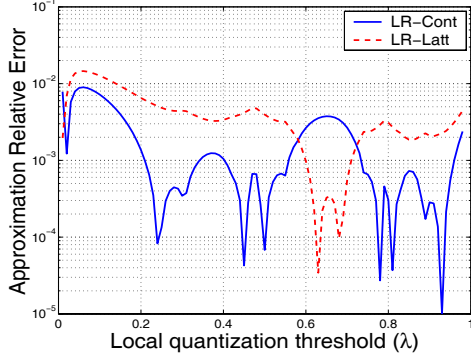
**Fig. 3**: Comparison between the continuous and the lattice versions of the saddlepoint approximation.
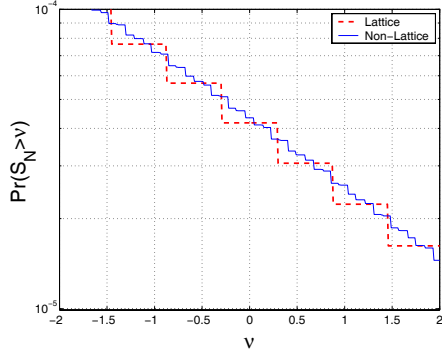


**Fig. 4**: Exact right-tail probability at different thresholds under lattice and non-lattice conditions.

error of the LR-Cont and LR-Latt approximations (relative error $= |(P_e)_{\text{exact}} - (P_e)_{\text{approx}}|/(P_e)_{\text{exact}}$). The figure demonstrates the advantage of the LR-Cont approximation for most of the range except around $\lambda = 0.65$. To explain this, notice that at

$$\lambda_{\text{Latt}} = \sqrt{\frac{\sigma^2}{2}} \log\left[\frac{1}{2}\left(-1 + \sqrt{1 + 8e^{\mu\sqrt{2/\sigma^2}}}\right)\right] \simeq 0.651 \quad (9)$$

the values of $L_m$ belong to a lattice with span $\beta = \sqrt{\frac{8}{\sigma^2}}\lambda_{\text{Latt}} \simeq 0.5822$. This is further illustrated in Fig. 4, where the exact probability $\Pr(S_N > \nu)$ is plotted when the local quantization threshold $\lambda = \lambda_{\text{Latt}} \simeq 0.651$ (corresponding to the lattice case) and when the quantization threshold is set to an arbitrary value $\lambda = 0.5$. The network size is fixed at $N = 101$ while the fusion threshold $\nu$ is varied from $-2$ to $2$ (i.e. changing the priors $\pi_0$ and $\pi_1$). The plot clearly illustrates the regular wide jumps for the lattice case. In contrast, when $\lambda = 0.5$, the jumps become irregular and closely spaced. As the number of sensors $N$ is increased the jumps become even closer (not shown here due to space limitation).

**Example 3** We illustrate the application of the proposed approximation to the problem of global optimization of sensor networks. Using the same setup as in Example 2, where the noise is assumed to be Laplacian, we find the optimum local quantization threshold $\boldsymbol{\lambda}_{\text{optim}} = (-\lambda_{\text{optim}}, 0, \lambda_{\text{optim}})$ that minimizes the global detection error. The proposed Lugannani-Rice approximation (LR-Cont) of the error performance is embedded in an algorithm to solve this optimization problem numerically for a wide range of operating SNR and the results are shown in Fig. 5. Interestingly, the optimum thresholds $\lambda_{\text{optim}}$ coincide with the uniform
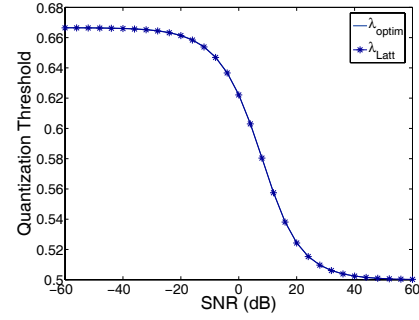


**Fig. 5**: Optimum thresholds $\lambda_{\text{optim}}$ obtained through numerical optimization when the noise is Laplacian. These coincide with the value of $\lambda_{\text{Latt}}$ in (9).

quantization thresholds $\lambda_{\text{Latt}}$ obtained from (9) ($\lambda_{\text{Latt}}$ corresponds to uniform quantization of the LLR). This shows that, for this example, which involves Laplacian noise, the optimal thresholds lead to a lattice-valued LLR $\ell$ and, hence, the simple majority rule in (7) is optimal. This is not necessarily true for other distributions. Results (not shown here) show that when the noise is Gaussian, the optimal thresholds lead to a non-lattice $\ell$, in which case the optimal fusion rule (1) may not be equivalent to the simple majority rule in (7) except for some special cases such as the small networks considered in [2]. Note, however, that the techniques proposed in this paper can be applied with high accuracy and low complexity regardless of whether $\ell$ is lattice or not. To verify the results in Fig. 5, we prove elsewhere that, asymptotically when $N \to \infty$ and SNR$\to -\infty$ dB, the optimum threshold is $\lambda = 2\mu/3$; and when SNR$\to +\infty$ the optimum threshold approaches $\mu/2$. These agree with the limits of (9) as SNR$\to \pm\infty$ and are close to the numerical results in Fig. 5 (we assumed that $\mu = 1$).

**Conclusion** The cooperation of a network of distributed sensors has many potentials although, individually, each sensor may not accomplish much. It is impressive to see how the collaboration of 100 sensors caused the error rate to drop down to $10^{-4}$, even under harsh conditions with an SNR of only $-10$ dB in the last example. The approximation proposed here simplifies the task of global optimization of such powerful networks without jeopardizing the performance. It enable us to optimize various networks ranging from the very small, with few number of sensors, to the very large, with a number of sensors in the thousands or even millions.

## 7. REFERENCES

[1] Jens L. Jensen, *Saddlepoint Approximations*, Oxford University Press, New York, NY, 1995.

[2] Saeed A. Aldosari and José M. F. Moura, "Fusion in sensor networks with communication constraints," in *Proc. of The IEEE/ACM Symposium on Information Processing in Sensor Networks (IPSN)*, Apr. 2004, pp. 108–115.

[3] John N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr., Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[4] Pramod K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, 1996.

[5] James G. Booth, Peter Hall, and Andrew T. Wood, "On the validity of Edgeworth and saddlepoint approximations," *Journal of Multivariate Analysis*, vol. 51, pp. 121–138, Oct. 1994.