Optimized Data Fusion in Bandwidth and Energy Constrained Sensor Networks

Xianren Wu and *Zhi Tian*
Department of Electrical and Computer Engineering
Michigan Technological University, Houghton, MI 49931, USA

***Abstract*— This paper considers the problem of decentralized data fusion (DDF) for large wireless sensor networks with stringent bandwidth requirements. To reduce the power and bandwidth costs of wireless transmissions, each sensor node is confined to quantize its sensing data and send 1-bit information only. Under this setting, we derive the maximum likelihood (ML) data fusion rule for decentralized parameter estimation, and analyze its Cramer-Rao lower bound (CRLB) of the fusion performance in the sense of mean square distortion. Depending on the underlying noise characteristics, our 1-bit DDF scheme can achieve estimation performance competitive to or even surprisingly better than that of centralized fusion over unquantized data. There is considerable saving in communication costs, which in turn reduces network energy consumption. Furthermore, we investigate network optimization, for which a worst-case robust design methodology is adopted to formulate a well-behaved min/max optimization problem. From the information processing viewpoint, the resulting optimized network offers robust fusion performance at minimal costs of communication resources.**

Keywords: decentralized data fusion, distributed signal processing, CRLB, network optimization, sensor networks

I. Introduction

One fundamental problem in a wireless sensor network (WSN) is data fusion, where observed data from sensors need to be aggregated to collectively reach an estimate of the underlying physical process of interest, subject to certain time delay constraints [1]. In order to improve the network lifetime and survivability, it is essential to conserve power during both data processing and communications. Recent studies show that communication tasks consume a large portion of the total energy needed for the overall network operation [2]. To strike the desired tradeoff in performance and power consumption, a key problem is to develop effective data fusion techniques under constraints imposed by limited communications.

To save power and bandwidth during data transmissions, sensed observations are often compressed and/or quantized locally at sensor nodes, and sent through lossy wireless channels to a fusion center/node for information aggregation [1]. Inevitably, distortions due to quantization, compression and transmission will affect data fusion. From information theoretic perspectives, [3], [4], [5] link the data fusion problem to Wyner-Ziv theory of source coding with decoder side information. Rate-distortion theory is then applied to derive the admissible distortion regions. These results provide important insights on where the distortion could be, but implementation of the coding strategy might not be practical under the storage and delay constraints. Centralized data fusion is addressed in [6], [7] by optimally designing local quantizers based on known joint distribution of sensor data. When the joint distribution is unavailable, it can be estimated via training [8].

Alternative to the unconstrained fusion approach where centralized fusion of unquantized data is performed, decentralized data fusion (DDF) schemes using 1-bit per sensor node have been proposed recently, which result in only linear performance penalty to optimal unconstrained fusion [9], [10], [11]. Sensors are organized into multiple groups, and each sensor only sends 1-bit local decision to the fusion center. Sensors within a group use the same local decision rule to process their observations, but the decision rules differ among groups. When the noise probability density function (pdf) is unknown but bounded, a universal (pdf-unware) but biased scheme entails performance penalty four times that of the optimal unconstrained one [9]. When the noise pdf is known and Gaussian, an optimal 1-bit fusion scheme can reduce the penalty to as small as $\pi/2$-fold [10]. The unknown noise pdf case is treated by a practical yet biased fusion scheme [11].

Focusing on DDF with 1-bit decision and transmission per sensor node, we consider in this paper the general case of optimal ML data fusion when the noise pdf is arbitrary (as oppose to Gaussian only). A new distributed multi-decision fusion scheme is developed, where multiple local decision rules are employed in the network. The fusion performance of the proposed ML scheme is analyzed by quantifying the Cramer-Rao lower bound (CRLB) of the mean square distortion in the fused data. Our 1-bit DDF scheme has fusion performance competitive to or even surprisingly better than the unconstrained fusion scheme, depending on the underlying noise characteristics. Meanwhile, 1-bit DDF considerably reduces the required bandwidth, which in turn saves energy.

Having obtained the optimal distributed data fusion scheme, we then turn to the problem of network optimization, where the goal is to optimize the grouping of sensor nodes and optimally design the local decision rules. We adopt a worst-case design methodology to formulate the network resource allocation problem as a min/max optimization problem that is readily solvable by software package [12]. The resulting optimized network satisfies robust performance requirements at minimal resource consumption.

This paper is organized as follows. Problem statement and formulation are introduced in Section II. Section III derives the optimal DDF data fusion rule and its analytic performance. Network optimization is covered in Section IV, followed by design examples in Section V and summary in Section VI.

II. Signal Model & Problem Formulation

Consider a WSN depicted in Fig. 1. A set of N sensors have been deployed to make observations of an unknown physical phenomenon, denoted by a scalar θ. These observations are:

$$x_k = \theta + n_k, \quad k = 1, 2, \ldots, N \tag{1}$$

where n_k denotes *i.i.d.* zero-mean random noise with pdf $p_k(u)$. In a homogeneous sensing environment, one has $p(u) := p_k(u)$, $\forall k$, where $p(u)$ is the common noise pdf.

Under bandwidth constraints, sensors perform local quantization or compression on $\{x_k\}$ before sending them to the fusion center. Suppose that I local quantization or compression functions are available at sensor nodes, denoted by the function set $\{f_i(x), 1 \leq i \leq I\}$. Each function maps a continuous observation x_k to a certain value selected in a discrete set, each represented by a finite number of bits N_b. In the case of minimal bandwidth requirement ($N_b = 1$), $f_i(x)$ becomes a binary decision function on sensor observations. These decision functions $\{f_i(x)\}$ are assumed to take the form:

$$f_i(x) = \begin{cases} 1, & x \leq \tau_i \\ 0, & x > \tau_i \end{cases} \qquad 1 \leq i \leq I \tag{2}$$

where τ_i is a scalar constant acting as a threshold in the corresponding decision function $f_i(x)$.

Each of the N sensors selects one of the I decision functions to implement locally, and is accordingly organized into one of the I sensor groups. Suppose that the i-th group uses $f_i(x)$ as its decision function, and the corresponding thresholds are arranged in an increasing order without loss of generality, i.e., $\tau_1 < \tau_2 < \cdots < \tau_I$. Representing the number of sensors in the i-th group by N_i, we define an I-element grouping vector $\mathbf{p} = (p_1, \ldots, p_I)$ where $p_i := N_i/N$. The local decisions $\{b_k\}$ of all N sensors are organized into a decision vector $\mathbf{b} := (b_1, \ldots, b_N)$. The following equalities arise:

$$\begin{aligned} N &= \textstyle\sum_{i=1}^{I} N_i; && (3)\\ 1 &= \textstyle\sum_{i=1}^{I} p_i, \quad p_i \geq 0, \quad 1 \leq i \leq I; && (4)\\ b_k &\in \{0, 1\}, \qquad 1 \leq k \leq N. && (5) \end{aligned}$$

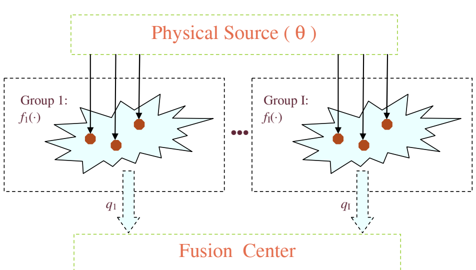

Fig. 1. Structure of a sensor network configured for data fusion.

Let $G(\cdot)$ be the global fusion function and $\hat{\theta}$ be the aggregate estimate of θ at the output of the fusion processor, via $\hat{\theta} = G(\mathbf{b})$. The fusion performance is measured by the mean square distortion $\epsilon(\theta) := E\{(\hat{\theta} - \theta)^2\}$. Our data fusion problem can be formulated as finding the fusion function $G(\cdot)$, the decision set $\{f_i(x)\}$ and the grouping vector $\mathbf{p}$ in order to minimize $\epsilon(\theta)$. Mathematically, this problem can be expressed as:

$$\min_{G, \{f_i\}, \mathbf{p}} \quad \epsilon(\theta). \tag{6}$$

III. DDF: Optimizing Global Fusion Function

We first consider the 1-bit decentralized data fusion (DDF) problem with known noise pdf $p(u)$. The unknown pdf case involves pdf estimation, which is not discussed for space limit.

A. Unconstrained Data Fusion (UDF)

In the absence of bandwidth constraints, sensor observations could be transmitted to the fusion center *lossless*. In this case, the global fusion function $G^*(\cdot)$ with respect to the sensor observation vector $\mathbf{x} := (x_1, \ldots, x_N)$ can use the best linear unbiased estimator (BLUE) for UDF as follows [13]:

$$\hat{\theta} = G^*(\mathbf{x}) = N^{-1} \textstyle\sum_{k=1}^{N} x_k = \theta + N^{-1} \sum_{k=1}^{N} n_k. \tag{7}$$

The BLUE G^* is independent of the underlying noise characteristics, and becomes optimal in the ML sense for Gaussian noise. Its associated mean square distortion (MSD) $\epsilon^*(\theta)$ is:

$$\epsilon^*(\theta) = \sigma^2 / N \tag{8}$$

where $\sigma^2 = \int u^2 p(u) du$ is the noise variance. The MSD of UDF serves as the performance reference to our 1-bit DDF.

B. DDF: Optimizing Global Fusion

Under bandwidth constraints, the fusion center only receives a discretized version of the sensor observation vector $\mathbf{x}$ that is used for data fusion. In the extreme case (1-bit per sensor), only the decision vector $\mathbf{b}$ is available to the fusion center. For the time being, we suppose that that N sensors have been classified into I groups based on known local decision functions $\{f_i\}$ and the grouping vector $\mathbf{p}$. We thus focus on the problem of designing the global fusion $G(\mathbf{b})$ with the goal of minimizing $\epsilon(\theta)$. The problem of designing I, $\{f_i\}$ and $\mathbf{p}$ will be investigated for network optimization in Section IV.

Consider the k-th group comprised of N_k sensor nodes. Let us focus on one of these N_k sensor nodes, say $N_{k,j}, 1 \leq j \leq N_k$, and define $b_{k,j}$ as its decision. From the decision rule in (2), the probability of $b_{k,j}$ taking 1 or 0 is given by

$$\begin{aligned} P(b_{k,j} = 1) &= \textstyle\int_{-\infty}^{\tau_k - \theta} p(u) du := F(\tau_k - \theta) \\ P(b_{k,j} = 0) &= 1 - F(\tau_k - \theta) \end{aligned} \tag{9}$$

where $F(x) = \int_{\infty}^{x} p(u) du$ is the corresponding cumulative distribution function (CDF) of the noise pdf $p(u)$. Clearly, $b_{k,j}$ is a binary variable with its mean and variance given by $E\{b_{k,j}\} = F(\tau_k - \theta)$ and $var\{b_{k,j}\} = F(\tau_k - \theta) - F(\tau_k - \theta)^2$.

By averaging decisions from these N_k sensor nodes, the aggregate observation $\hat{q}_k$ of k-th group can be written as:

$$\hat{q}_k = N_k^{-1} \textstyle\sum_{j=1}^{N_k} b_{k,j}. \tag{10}$$

When N_k is sufficiently large to invoke the central limit theorem (CLT), $\hat{q}_k$ can be modeled as a Gaussian random variable with mean q_k and variance σ_k^2. For simplicity, σ_k^2 is approximated as constant and measured from the environment. The mean value q_k is given by

$$q_k = N_k^{-1} \textstyle\sum_{j=1}^{N_k} E\{b_{k,j}\} = F(\tau_k - \theta). \tag{11}$$

Here $\hat{q}_k$ can be viewed as a noisy observation of $F(x - \theta)$ at the sample point τ_k that belongs to the vector $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_I)$.

The conditional probability of observing $\hat{q}$ given $\boldsymbol{\tau}$ and θ is

$$J(\theta) = p(\hat{\mathbf{q}}|\theta, \boldsymbol{\tau}) = \prod_{k=1}^{I} p(\hat{q}_k|\theta, \tau_k) = \prod_{k=1}^{I} \mathcal{N}(\hat{q}_k; q_k, \sigma_k^2). \tag{12}$$

The optimal global fusion function $G(\cdot)$ should find the $\hat{\theta}$ that maximizes this conditional probability, expressed as follows:

$$\begin{aligned} \hat{\theta} &= G(\mathbf{b}) = G(\hat{\mathbf{q}}) = \arg\max_{\theta} \ \log J(\theta) \\ &= \arg\min_{\theta} \ \textstyle\sum_{k=1}^{I} (\hat{q}_k - q_k(\theta))^2/\sigma_k^2. \end{aligned} \tag{13}$$

The optimal $\hat{\theta}$ can be derived from (13) by forcing the first-order derivative of $\log J(\hat{\theta})$ to be zero, which leads to

$$\textstyle\sum_{k=1}^{I} (\hat{q}_k - F(\tau_k - x))F'(\tau_k - x)/\sigma_k^2 \Big|_{x=\hat{\theta}} = 0. \tag{14}$$

There is no closed-form solution to (14); nevertheless, $\hat{\theta}$ can be solved by efficient numerical methods, either using a Newton descent-search algorithm on (13) or applying a bi-sectional search on (14). These search algorithms have well-behaved convergence, because we can prove the formulation in (13) asymptotically (in N) reaches a unique global optimum point.

Without delving into the detailed derivations, we can prove that the CRLB of $\hat{\theta}$ for the unbiased ML estimator $G(\cdot)$ is:

$$\epsilon(\theta) = E(\theta - \hat{\theta})^2 \geq \frac{1}{N \sum_{k=1}^{I} p_k \mu_\theta(\tau_k)} \tag{15}$$

where $\mu_\theta(x) = \frac{p(x-\theta)^2}{F(x-\theta)(1-F(x-\theta))}$.

To gain more insight on the CRLB, let $p_U(u)$ be a noise pdf with normalized unit-variance, and $F_U(u)$ be the corresponding CDF, that is, $p(u) = \frac{1}{\sigma} p_U(\frac{u}{\sigma})$ and $F(u) = F_U(\frac{u}{\sigma})$. Correspondingly, the CRLB can be rewritten as

$$\epsilon(\theta) \geq K(\theta) \cdot \frac{\sigma^2}{N}, \qquad K(\theta) := \frac{1}{\sum_{k=1}^{N} p_k \mu\left(\frac{\tau_k - \theta}{\sigma}\right)} \tag{16}$$

where $\mu(x) := \frac{p_U(x)^2}{F_U(x)(1-F_U(x))}$ determines a linear performance penalty factor $K(\theta)$ with reference to the UDF result $\epsilon^*(\theta)$ in (8). Fig. 2 depicts $\mu(x)$ for unit-variance Gaussian and Laplacian noise pdfs. It can be observed that $\mu(x)$ peaks at the origin, indicating that the MSD is minimized when $\tau_1 = \cdots = \tau_I = \theta$. As such, the 1-bit DDF suffers a moderate linear performance penalty of $K(\theta) = \pi/2$ for the Gaussian noise, but exhibits better MSD for the Laplacian noise since $\mu(0) = 0.5$. This is because BLUE is ML optimal only for Gaussian noise.

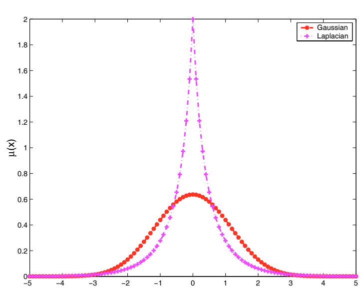

Fig. 2. CRLB-related $\mu(x)$ of unit-variance Gaussian and Laplacian noises.

IV. DDF: Network Optimization

The CLRB in (15) suggests that thresholds $\boldsymbol{\tau}$ in local decision functions need to be set exactly at θ to minimize the performance penalty in the 1-bit DDF scheme. Since the source variable θ may vary over time due to environmental changes, these thresholds need to be updated to be close to θ for satisfactory performance. However, updating the thresholds leads to communication overhead that may offset the low communication benefits of the 1-bit DDF scheme.

Assuming the dynamic range of the source variable θ is known a prior, we present a design strategy that provides robust fusion performance for any θ within the known range, with no overhead in updating the thresholds. Noting that the CRLB in (16) only depends on the relative distance between τ and θ, we assume without loss of generality that the dynamic range can be expressed as $\theta \in (-\Omega, \Omega), \Omega > 0$. Our robust design boils down to optimization of the network so that the least number N of sensor nodes is needed to provide fusion performance no worse than a desired level ϵ_0, i.e.,

$$\min_{\{f_i\},\mathbf{p}} \quad N \qquad s.t. \ \ \epsilon(\theta) \leq \epsilon_0, \ \forall \ \theta \in (-\Omega, \Omega). \tag{17}$$

This formulation provide guaranteed performance with robustness to any change of θ within a range. Minimizing N is also useful in reducing the load at the medium access control (MAC) layer to avoid packet collision when scheduling transmissions from a large number of sensors. It is reasonable to substitute the unknown $\epsilon(\theta)$ in (17) by its CRLB in (16), since $\epsilon(\theta)$ of our optimal 1-bit DDF will approach the CRLB when N is sufficiently large, which is typically the case for a large-size dense WSN. With this approximation $\epsilon(\theta) = K(\theta) \cdot \sigma^2/N$, it can be deduced from (17) that

$$N \geq C \cdot K(\theta), \qquad \text{where} \ \ C = \sigma^2/\epsilon_0 > 0. \tag{18}$$

Clearly, the design objective in (17) can be transformed to

$$\min_{\{f_i\},\mathbf{p}} \ \max_{\theta} \ K(\theta). \tag{19}$$

Design A [Uniform Network Distribution]: Suppose the N sensors are uniformly organized into I groups and the thresholds in local decision functions are evenly distributed within the dynamic range of the source variable θ, i.e., $p_i = 1/I, \ \forall i$, and $\tau_k = -\Omega + 2\Omega * \frac{k}{I+1}$, $1 \leq k \leq I$. For every $\theta \in (-\Omega, \Omega)$, there is at least one $\tau_k, k \in \{1, \ldots, I\}$ that satisfies $|\tau_k - \theta| \leq \frac{2\Omega}{k+1}$. It can be deduced that

$$K(\theta) \ \ \leq \ \ I/\mu\left(\frac{1}{I+1} \cdot \frac{2\Omega}{\sigma}\right). \tag{20}$$

From (20), a numerical bi-sectional search on I yields the optimal group number I_{opt}, which depends on the desired ϵ_0.

Design B [Optimized Network Distribution]: The network design can be improved by adopting a non-uniform strategy. Since $K(\theta) \geq 0$, we reformulate our optimization problem as:

$$\min_{\{f_i\},\mathbf{p}} \ \max_{\theta} \quad K(\theta) \qquad s.t. \ \ \textstyle\sum_{k=1}^{I} p_i = 1, \quad \tau_1 < \cdots < \tau_I. \tag{21}$$

This is a standard min/max optimization problem that can be solved efficiently using off-the-shelf software, e.g., [12].

V. Design Examples & Simulation Results

Simulations are conducted to evaluate the performance of the proposed 1-bit DDF scheme and compare it with the UDF scheme. In the first design example, all sensor nodes use a common local decision function. The threshold associated with this local decision function is periodically updated to be close to the underlying physical phenomenon θ. Fig. 3 depicts the MSD performance of the proposed fusion rule in (13), along with the CRLB and the performance of the UDF. Both unit-variance Gaussian and Laplacian noise pdfs are considered.

It shows that when N is sufficient large, the MSD $\epsilon(\theta)$ of 1-bit DDF matches the CRLB very well, which confirms the ML optimality of our design and justifies the use of the CRLB value for network optimization in (17). Furthermore, Fig. 3 confirms that the 1-bit DDF has estimation performance linear to that of the UDF. Depending on the underlying noise characteristics, the linear performance penalty factor is $K(\theta) = \pi/2$ for Gaussian noise and $K(\theta) < 1$ for Laplacian noise, the latter of which corresponds to better fusion performance than BLUE-based UDF.

The second simulation example examines network optimization under both uniform and non-uniform deployment strategies. In Fig. 4, the dynamic range of θ is $(-2, 2)$, the group number is $I = 2$, and both Gaussian and Laplacian noise pdfs of unit variance are evaluated. The non-uniform deployment strategy demonstrates better performance than the uniform one, at the expense of higher complexity needed in determining the optimal thresholds. The robust deployment for the 1-bit DDF avoids the overhead of updating thresholds by settling for (non-optimal) guaranteed performance at a prescribed level ϵ_0. In contrast, the UDF scheme is independent of the dynamic range of the underlying θ.

VI. Conclusions

For decentralized data fusion using 1-bit per sensor node, we derived in this paper the optimal ML data fusion rule and its CRLB performance under known noise pdf. Both theoretical and simulation results show that the proposed scheme can achieve fusion performance comparable to or better than unconstrained data fusion, at minimal bandwidth consumption. Network optimization is also considered, which is transformed as a robust optimization problem through min/max formulation. Both uniform and non-uniform deployment strategies are considered and optimized to minimize network resource consumption. The network optimization solutions minimize the number of sensor nodes needed, which is also useful in reducing the load at the MAC layer to avoid packet collision during sensor scheduling. These results are discussed for the known noise pdf case, and will be instrumental to solving the more challenging case of unknown noise characteristic.

References

[1] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: part II - advanced topics," *Proceedings of the IEEE*, pp. 64-79, Jan. 1997.

[2] V. Shnayder, M. Hempstead, B.-R Chen and G. Werner-Allen and M. Welsh, "Simulating the Power Consumption of Large-Scale Sensor Network Applications," *Proc. of ACM Conf. on Embedded Networked Sensor Systems (SenSys'04)*, Baltimore, MD, November 2004.

[3] S. C. Draper and G. W. Wornel, "Side Information Aware Coding Strategies for Sensor Networks," *IEEE Journal on Selected Areas In Communications*, vol. 22, no. 6, pp. 966 - 976, August 2004.

[4] J. Chen, X. Zhang, T. Berger and S. B. Wicker, "An Upper Bound On the Sum-Rate Distortion Function and Its Corresponding Rate Allocation Schemes for the CEO Problem," *IEEE Journal on Selected Areas In Communications*, vol. 22, no. 6, pp. 977 - 987, August 2004.

[5] P. Ishwar, R. Puri, S. S. Pradhan and K. Ramchandran, "On Rate-constrained Estimation in Unreliable Sensor Networks," *Proc. of IPSN Conf.*, pp. 178 - 192, 2003.

[6] J. A. Gubner, " Distributed Estimation and Quantization," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1456 - 1459, July 1993.

[7] W.-M Lam and A. R. Reibman, "Design of Quantizers for Decentralized Estimation Systems," *IEEE Trans. on Communications*, vol. 41, no. 11, pp. 1602 - 1605, November 1993.

[8] V. Megalooikonomou and Y. Yesha, "Quantizer Design for Distributed Estimation with Communication Constraints and Unknown Observation Statistics," *IEEE Trans. on Communications*, vol. 48, no. 2, pp. 181 - 184, February 2000.

[9] Z.-Q. Luo, "Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network," *IEEE Trans. on Info. Theory*, to appear.

[10] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part I: Gaussian Case," *IEEE Trans. on Signal Processing*, to appear.

[11] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part II: Unknown pdf," *IEEE Trans. on Signal Processing*, to appear.

[12] The Mathworks Inc., "MATLAB Optimization Toolbox," available at http://www.mathworks.com.

[13] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory," Prentice Hall, 1993.

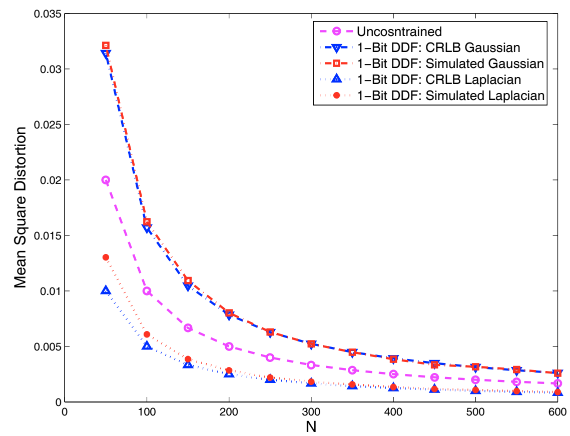

Fig. 3. Fusion performance for various optimal global fusion functions.

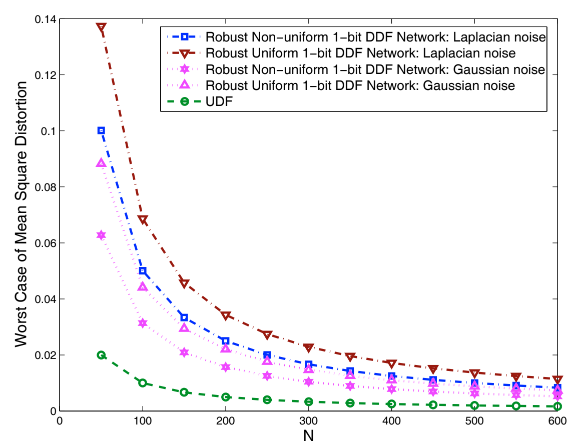

Fig. 4. Fusion performance for various network optimization strategies.