This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2008 proceedings.

Asynchronous Sampling Benefits Wireless Sensor Networks

Jing Wang, Yonghe Liu, and Sajal K. Das
Center for Research in Wireless Mobility and Networking(CReWMaN)
Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, Texas 76019
Email: {jwang,yonghe,das}@cse.uta.edu

***Abstract*— Intensive research has focused on redundance reduction in wireless sensor networks among sensory data due to the spatial and temporal correlation embedded therein. In this paper, we propose a novel approach termed asynchronous sampling that complements existing study. The key idea of asynchronous sampling is to spread the sampling times of the sensor nodes over the time line instead of performing them in a synchronous manner. Compared with existing strategies, asynchronous sampling introduces another dimension for optimization, without additional computation or communication overhead on sensor nodes. Theoretically, we show that asynchronous sampling benefits sensor networks through increased entropy of the sensory data or reduced reconstruction distortion. Furthermore, we formulate the optimal asynchronous sampling problem for determining the time shifts among the nodes. A heuristic solution, termed O-ASYN, is presented that uses local optimum search to approximate the global optimal solution. Simulation results based on simulated data and real experimental data both demonstrate the entropy increases.**

I. Introduction

Facing the severe resource limitations in wireless sensor networks in particular that of energy, an extensive body of research has centered on extending the network life time of wireless sensor networks from difference perspectives. Among them a set of papers have investigated rate allocation in such networks. Given the resource constraint, the focus there is to extend the network lifetime and network capacity subject to the computation cost to process the sensor data and the communication cost for data transmission, which is often determined by the sampling rates (data flow rates) of the sensor nodes [1–3].

A key finding there is that lowering the sampling (data flow) rates of sensor nodes can significantly prolong the lifetime of the network and improve transport reliability of multi-hop communication. For instance, it's reported in [4] that lifetime of the sensor network can increase extraordinarily from one month to more than 18 months by lowering the sampling rate from producing a sample every 30 seconds to every 10 minutes. Experiments with TinyOS platform conducted in [5] show that packet delivery ratio increases from around 55% to 81% when source packet generation rate drops from 7.69/s to 4/s.

However, when sampling rate is reduced in favor of longer lifetime or reliable packet delivery, quality of the information collected from sensor nodes will deteriorate accordingly. This may imply insufficient information about the physical field of interest and in turn low estimation accuracy or even failure of event detection.

In this paper, we propose an asynchronous sampling strategy to simultaneously achieve low sampling rate (hence the resulting benefits) and high information quality . By allowing sensor nodes to sample correlated physical variables asynchronously, entropy of the sensor data increases, which in turn counterbalances the loss of lowering sampling rate. Our simulation results show that the lower the sampling rate, the more benefits can be introduced by asynchronous sampling in terms of alleviating information deterioration.

The rest of this paper is organized as follows. Related work is presented in Section II firstly. Correlation models and examples are introduced in Section III. The optimization problem is put forward in Section IV using correlation models. Then the asynchronous sampling strategy is introduced in Section V, which is accompanied by the simulation on the strategies shown in Section VI. At last, Section VII concludes the paper and discusses future work.

II. Related Work

Due to the nature of the physical phenomena observed by wireless sensor networks, sensory data are often correlated [6, 7]. As achieving energy-efficiency is one of the primary goals, data compression or aggregation techniques are often adopted to reduce redundancy of sensor data and hence computing or communication cost [8–10]. Sensor node selection [11, 12] and work scheduling of sensor nodes [13] lead to fewer number of nodes involved at a certain time point for data gathering and hence can conserve energy as well. Often in these techniques, spatial correlation plays an important role in reshaping the aggregation and compression algorithms with the knowledge of correlation among the neighboring sensor nodes.

Optimal spatial sampling aims at selecting sensor nodes that contribute most to retrieved information about the physical phenomena while satisfying energy consumption constraints [14]. For example, adaptive spatial sampling discussed in [15] is able to reduce communication cost while maintaining high accuracy due to its three-phase process to trigger sensor nodes based on the current fusion result.

In this paper, we show in addition to above approaches, asynchronous sampling introduces another dimension for energy conservation with little computation and communication overhead. Our key idea is to assign different sampling shifts to individual sensor node and hence create an asynchronous sampling strategy in the network. This is in stark contrast with existing approaches where samplings are performed mostly in an synchronized fashion.

III. Temporal-Spatial Correlation Models

The key benefit of asynchronous sampling is redundancy reduction of sensory data, which can in turn be exploited to maximize entropy of the information or minimize energy consumption (through reduction of sampling rate). Before detailing our asynchronous sampling strategies, we will first discuss the correlation model we use in this paper.

A. Temporal-Spatial Correlation

We consider a dense senor network monitoring a physical process such as wind speed or temperature field. The sensory data gathered by a sensor node is composed of the measurement value and a noise. Assume that the location of the sensor node i is denoted by (x_i, y_i, z_i), a sensory sample S_i at time t can be expressed as

$$S_i(t) = M(x_i, y_i, z_i, t) + N(x_i, y_i, z_i, t), \tag{1}$$

where M represents the measurement (true) value determined by (x_i, y_i, z_i) and t, and N is the noise introduced by the environment or the sampling process.

Assume that the sensory data given in (1) are Joint Gaussion Random Variables (JGRVs) with zero mean and σ_M^2 variances, and noise N_i is i.i.d Gaussian random variable with zero mean and σ_N^2 variances. Then the spatial correlation between sensor node i and j can be expressed as

$$\rho_s(i,j) = \frac{E[S_i S_j]}{\sigma_M^2 + \sigma_N^2}. \tag{2}$$

As a commonly employed model [16], the spatial correlation is assumed to be inversely proportional to the distance between the nodes:

$$\rho_{i,j}^{spatial} = e^{-\alpha d_{i,j}} = e^{-(\alpha(\sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2})} \tag{3}$$

where $\alpha > 0$ denotes a constant for spatial correlation intensity and $d_{i,j}$ is the distance between node i and j.

Temporal correlation often denotes the correlation between data sampled at different time points. Similarly, for a wide-sense stationary process [17] as Gaussian random process, the temporal correlation can be expressed as

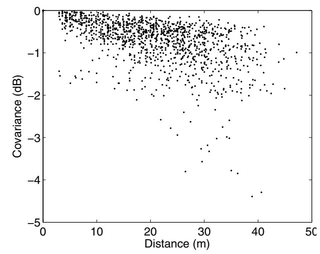

Fig. 1: Coefficients and distances

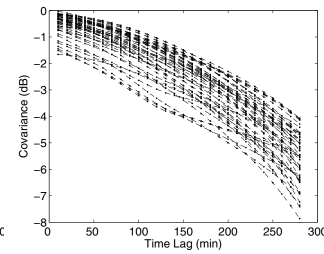

Fig. 2: Coefficients and time lags

$$\rho_{i,j}^{temporal} = e^{-\beta \tau_{i,j}} \tag{4}$$

where $\tau_{i,j}$ is the distance between the sampling time of nodes i and j, i.e., $\tau_{i,j} = |t_j - t_i|$, and β is the constant measuring the temporal correlation intensity. Combining both spatial and temporal correlations, we can define

$$\rho_{i,j} = e^{-(\alpha d_{i,j} + \beta \tau_{i,j})} \tag{5}$$

as the correlation of sensory data between nodes i and j.

B. Example

Here, we present an example to verify the above exponential correlation model. We use a set of experimental data from Intel Berkeley Lab [18] for this purpose. This set of data represents temperatures measured at different locations in a lab space. Given the temperature measurements and locations of sensor nodes, we calculate the covariances between any pair of sensor nodes. Fig. 1 shows the relationship between covariance coefficient and distance between two sensor nodes. From the figure, it is obvious that the experiment data don't fit the spatial correlation model in (3) although the covariance coefficients indeed tend to decrease when distance between the sensor nodes increases. For the sake of graph explicitness, we calculate the covariance coefficients using decibel to transform exponential relationship into linear relationship.

On the contrary, as shown in Fig. 2 the covariance coefficient of temperature measurements between a node and those taken after certain time period at another sensor node does exponentially decrease when this time lag increases. Here, each dashed line corresponds to a pair of sensor nodes. For each pair of sensor nodes, its covariance coefficient in decibel is inversely proportional to the time lag between the temperature measurements of the two nodes.

IV. Benefits of Asynchronous Sampling

Given the temporal-spatial correlation model discussed above, next we will show that asynchronous sampling strategy, if employed, can indeed increase the entropy of the data and reduce regression distortion as a result of increased entropy.

A. Asynchronous Sampling Increases Entropy

For tractability, we assume that sensory data are jointly Gaussian random variables. Entropy of the sensory data composed of $S_1, \cdots, S_n$ samples following this model can then be derived as

$$H = \frac{1}{2}\log(2\pi e)^n detK_n - \log\Delta. \quad (6)$$

Here, entry $k_{i,j}$ in the covariance matrix K_n corresponding to sensor data samples S_i and S_j can be expressed as

$$k_{i,j} = \begin{cases} \sigma_i^2 & i = j, \text{ and } i,j = 1,\cdots,n \\ \sigma_i\sigma_j & i \neq j, \text{ and } i,j = 1,\cdots,n \end{cases}$$

where σ_i and σ_j are the standard deviation of the samples S_i and S_j, respectively. $\log\Delta$ is a constant due to quantization. For the sake of simplicity, we normalize the covariance matrix K_n in order to get a correlation coefficient matrix A_n. Then the determinant of K_n can be derived as

$$detK_n = \prod_{i=1}^{n}\sigma_i^2 detA_n \quad (7)$$

With this, we have the following theorem.

Theorem 1.1: Entropy of sensor data $S_1, \cdots, S_n$ increases through asynchronous sampling.

Proof: Suppose that A_n is constructed from synchronous sampling. Now if we shift the sampling sequence of the jth sensor node t_j to $t_j + \tau$. Correlation between the jth node and the other sensor nodes $\rho_{i,j}$ changes to $\hat{\rho}_{i,j}$, where

$$\hat{\rho}_{i,j} = \rho_{i,j}e^{-\beta\tau_{i,j}} \qquad i,j = 1,...,n$$

Then, $\hat{A}_n$ in correspondence to asynchronous sampling is given by $\hat{A}_n = A_n \circ B_n$, where B_n is the sampling shift matrix whose entry is given by

$$b_{i,j} = \begin{cases} 1 & i = j, \text{ and } i,j = 1,\cdots,n \\ e^{-\beta\tau_{i,j}} & i \neq j, \text{ and } i,j = 1,\cdots,n \end{cases} \quad (8)$$

And $A_n \circ B_n$ is Hadamard product, which is the element-wise product of two matrixes. As A and B are correlation matrixes, they are positive definite or positive semidefinite. It is know that Hadamard product of two positive definite matrixes are also positive definite due to the closure property of positive definite matrix under Hadamard product. According to Oppenheim's Inequality [19], we have $det(A_n \circ B_n) \geq detA_n \prod_{i=1}^{n} b_{i,i} = detA_n$ (equality holds if and only if A_n is a diagonal matrix), which shows $detA_n < det\hat{A}_n$. Therefore, $detK_n < det\hat{K}_n$, which infers that $H < \hat{H} = \frac{1}{2}\log(2\pi e)^n det\hat{K}_n - \log\Delta$. ■

B. Asynchronous Sampling Reduces Distortion

While entropy provides an abstract quantification of the information amount embedded in the data, it is hard to picture its true impact in real applications. Here, we take one step further to show that asynchronous sampling can help an application improving regression of the physical process from the asynchronous data.

A key goal of deploying sensor networks is to reconstruct the physical field under measurement based on the gathered data. Instead of reconstruction, regression of the physical field is implemented when reconstruction is unreachable due to insufficient data. In this section, we show how asynchronous sampling can improve regression performance through an example of linear regression.

We consider a regression model for a physical process as shown below.

$$\widetilde{S}(x,y,z,t) = \sum_i w_i H_i(x,y,z,t)$$

where w_i is the ith weight, H_i is the ith basis function of regression. The optimal regression is achieved when $w = (H^TH)^{-1}H^TS$. Thus, the regress distortion is

$$D_r = E[(\widehat{S} - H(H^TH)^{-1}H^TS)^2]$$

where $\widehat{S}$ represents the physical value measured by the sensor nodes.

Let $Q = H(H^TH)^{-1}H^T$, we have

$$\widetilde{S}_j = \sum_{i=1}^{n} q_{i,j}S_i$$

Consequently,

$$\begin{aligned} D_r \quad &= \quad \sigma_S^2 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} q_{i,j}^2 S_i^2 \\ &+\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} q_{i,k}q_{j,k}\rho_{i,j}S_k^2 \\ &-\frac{2}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} q_{i,k}\rho_{S,i}\widehat{S}_k^2 \end{aligned} \quad (9)$$

If the sensory data are collected asynchronously, only $\rho_{i,j}$ will decrease due to non-zero $\tau_{i,j}$ introduced in the correlation model. As a result, we can conclude that asynchronous sampling can reduce regression distortion.

V. Asynchronous Sampling Strategies

As we have shown above, asynchronous sampling strategy increases entropy of the sensor data by introducing shifts in the sampling time points among different sensor nodes. In this section, we first formulate the optimization problem for determining the time shifts among the sensor nodes. As this optimization problem is NP-hard, we propose a heuristic algorithm termed as O-ASYN, which uses local optimum to approximate global optimum.

Without losing generality, we assume that the sampling time points of the sensor nodes is increasing along with their index, i.e., $t_1 \leq t_2 \leq \cdots \leq t_n$. We also assume that the times shifts among the nodes are

$$\tau_i = \begin{cases} t_{i+1} - t_i & i = 1,\cdots,n-1 \\ T + t_1 - t_i & i = n \end{cases}$$

where T is the sampling interval of the sensor nodes.

To best benefit from the asynchronous sampling strategy, our goal is to determine the best set of $\{\tau_i\}$ so that the entropy of the sensory data can be maximized. Formally, the goal is

$$\max logdet A_n \quad (10)$$
$$\text{subject to} \quad \sum_{k=1}^{n-1} \tau_k = T \text{ and } \tau_k \geq 0$$

The above optimization problem can be shown to be NP-hard as it's similar to the Assignment Problem with Extra Constraints (APEC), which is NP-hard [20].

As a result, we propose a recursive algorithm named O-ASYN to approximate the global optimum using local optimum. O-ASYN is described in Algorithm 1. In this algorithm, given the index of sensor nodes, we start from the local optimum of the optimization problem with 3 sensor nodes. As linearity of local optimum of k sensor nodes is maintained in searching for local optimum of $k+1$ sensor nodes, we are able to use Lagrange Multiplier recursively to obtain local optimum for n sensor nodes.

1. index the sensor nodes
   S: set of sensor nodes; V: set of sensor nodes with index;
   u: index of sensor nodes; V=$\emptyset$
   for i=1 to $|S|$
      $u_i = argmax_{u \subset S} Entropy(V \bigcup u)$
      V=$V \bigcup u_i$
      $S = S - u_i$
   end
2. find sampling shifts of the indexed nodes
   $\tau_1 : \frac{T}{2} + \frac{1}{2\beta}(\ln \rho_{1,2} - \ln \rho_{2,3})$; $\tau_2 : \frac{T}{2} + \frac{1}{2\beta}(\ln \rho_{2,3} - \ln \rho_{1,2})$
   are the local optimum of sampling shifts of u_1, u_2 and u_3
   for k=1 to n-3
      for j=1 to k+1
         $\tau_j = \tau_j (T - \tau_{k+2}) / (\sum_{i=1}^{k+1} \tau_i)$
      end
      $\tau_{k+2} = argmaxEntropy(\tau_1, \cdots, \tau_{k+1}, \tau_{k+2})$
         subject to $\sum_{i=1}^{k+1} \tau_i + \tau_{k+2} = T$ and $\tau_i \geq 0$
   end

Algorithm 1: O-ASYN algorithm

Besides O-ASYN, simple alternative asynchronous sampling strategies can be applied as well. For example, time shifts among sensor nodes can be randomly generated or equal among all. We denote them R-ASYN and and E-ASYN respectively. We will study the performance of the three algorithms in the next section.

VI. SIMULATIONS

In this section, we study the performance of O-ASYN using both simulated sensory data and real data obtained from experimental deployment of real sensor networks. With simulated sensor data, performance of O-ASYN in terms of increased determinant of covariance matrix is compared with those of two alternative asynchronous sampling

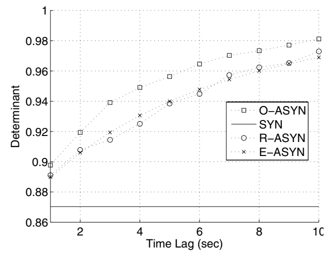

Fig. 3: Simulation on Maximum-Entropy Asynchronous Strategy

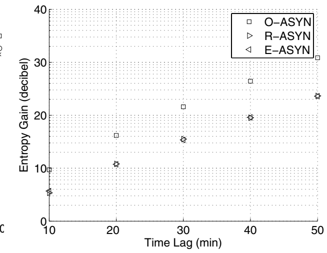

Fig. 4: Maximal Entropy Strategy Simulation

strategies, namely R-ASYN and E-ASYN (with random time shifts and equal time shifts among sensor nodes respectively). In addition, we verify the benefit of asynchronous sampling strategy by conducting asynchronous sampling strategies on real data set.

Simulated sensory data of 10 nodes are prepared following the correlation models given in Section III. We assume the sensor nodes are located in 3-dimension space. The spatial correlation of a pair of sensor nodes is determined by the distance between them. Locations of the sensor nodes are randomly distributed in a $10m \times 10m \times 10m$ space. The corresponding correlation matrix A_{10} is obtained by denoting $\alpha = 0.5$ and $\beta = 0.2$.

With simulated sensory data, we compute determinants of the covariance matrixes corresponding to synchronous samples and asynchronous samples respectively. Recall that determinant of covariance matrix is linearly proportional to entropy of the sensory data, results shown in Fig.3 shows increased entropy of asynchronous samples.

Besides simulated sensory data, real data set from Intel Berkeley Lab are adopted to verify the benefits of asynchronous sampling. The data set contains temperature measurements in a lab space for about one month. Among the 54 sensor nodes in the area, 50 nodes transmit valid sample series to the sink. Before computing correlation of the asynchronous samples, corrupted data are excluded during preprocess of the data set.

Similarly, performance comparisons among synchronous sampling and asynchronous sampling strategies are conducted when the maximum time shift varies from 30s to 150s. Simulation results on entropy gain are shown in Fig.4. Here, entropy gain is the ratio between entropy of asynchronous samples and that of synchronous samples in decibel. Not surprisingly, entropy gains of O-ASYN, R-ASYN and E-ASYN follow the similar trend as shown in Fig. 3. Notice that the larger the time shifts, the greater improvement is achieved by asynchronous sampling.

Furthermore, we compute the regression distortion of the temperature values given the sensor data sampled at reduced sampling rate. The regression results corresponding to synchronous samples and asynchronous samples are shown in Fig.5 and Fig.6 respectively, while the sensor

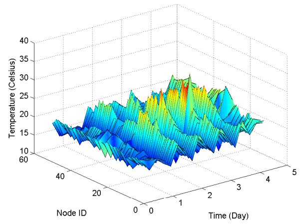

Fig. 5: Data sampled at reduced rate synchronously

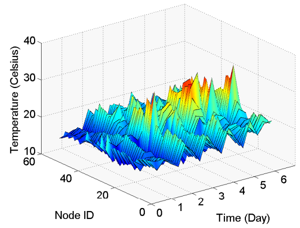

Fig. 6: Data sampled at reduced rate asynchronously

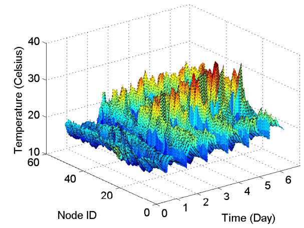

Fig. 7: Data sampled at original rate synchronously

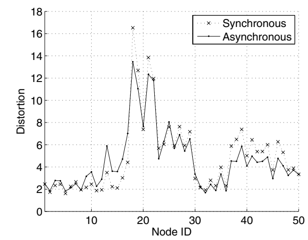

Fig. 8: Comparison of Regression Performance

data sampled at original sampling rate is show in Fig.7. Regression distortion of synchronous sampling and asynchronous sampling are compared in Fig.8. Evidently, the overall regression distortion of asynchronous sampling is much lower than that of synchronous sampling.

VII. Conclusion

In this paper, we propose asynchronous sampling where the sampling time of sensor nodes are shifted. Theoretically and experimentally, we show that entropy of sensory data increases as a result of correlation reduction due to asynchronous sampling. Furthermore, we model optimal asynchronous sampling as an optimization problem with an objective of entropy maximization with constraints on sampling time shifts. The heuristic solution we propose, termed O-ASYN, approximates the global optimum through recursively search of local optimum of the objective function. Compared with existing strategies focused on correlation reduction among sensory data, asynchronous sampling strategy introduces another dimension for optimization, without additional computation or communication overhead. Currently we are exploring the tradeoff between energy efficiency and reliability in the network based on the conclusions from this paper.

References

[1] X. Liu, Q. Wang, W. He, M. Caccamo, and L. Sha, "Optimal real-time sampling rate assignment for wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 2, no. 2, pp. 263–295, 2006.

[2] J. Zhu, K.-L. Hung, and B. Bensaou, "Tradeoff between network lifetime and fair rate allocation in wireless sensor networks with multi-path routing," in *MSWiM '06: Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, 2006, pp. 301–308.

[3] J. Paek and R. Govindan, "Rcrt: rate-controlled reliable transport for wireless sensor networks," in *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*, 2007, pp. 305–319.

[4] C. Guestrin, P. Bodi, R. Thibau, M. Paski, and S. Madde, "Distributed regression: an efficient framework for modeling sensor network data," in *IPSN '04: Proceedings of the third international symposium on Information processing in sensor networks*. New York, NY, USA: ACM Press, 2004, pp. 1–10.

[5] C.-Y. Wan, S. B. Eisenman, and A. T. Campbell, "Coda: congestion detection and avoidance in sensor networks," in *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*, 2003, pp. 266–279.

[6] J. Chou, D. Petrovic, and K. Ramchandran, "Tracking and exploiting correlations in dense sensor networks," in *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, USA, Nov. 2002, pp. 39–43.

[7] C. Tang and C. S. Raghavendra, "Correlation analysis and application in wireless microsensor networks," in *Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services(MobiQuitous'04)*, Boston, Massachusetts, USA, Aug. 2004, pp. 184–193.

[8] A. Jindal and K. Psounis, "Modeling spatially-correlated sensor network data," *ACM Transactions on Sensor Networks*, vol. V, pp. 1–35, 2006.

[9] A. T. Hoang and M. Motani, "Exploiting wireless broadcast in spatially correlated sensor networks," in *IEEE International Conference on Communications(ICC 2005)*, Seoul, Korea, May 2005, pp. 2807–2811.

[10] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," in *Third International Symposium on Information Processing in Sensor Networks(IPSN 2004)*, Berkeley, California, USA, Apr. 2004, pp. 28–35.

[11] A. Boulis, S. Ganeriwal, and M. B. Srivastava, "Aggregation in sensor networks: an energy-accuracy trade-off," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 317–331, 2003.

[12] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Third International Symposium on Information Processing in Sensor Networks(IPSN 2004)*, Berkeley, California, USA, Apr. 2004, pp. 20–27.

[13] K. Yuen, B. Liang, and B. Li, "Distributed data gathering in multi-sink sensor networks with correlated source," in *Proceedings of IFIP Networking 2006*, Coimbra, Portugal, May 2006, pp. 868–879.

[14] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in gaussian processes," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*. New York, NY, USA: ACM Press, 2005, pp. 265–272.

[15] R. Willett, A. Martin, and R. Nowak, "Backcasting: adaptive sampling for sensor networks," in *IPSN '04: Proceedings of the third international symposium on Information processing in sensor networks*. New York, NY, USA: ACM Press, 2004, pp. 124–133.

[16] P. J. Diggle, J. A. Tawn, and R. A. Moyeed, "Model-based geostatistics," *Applied Statistics*, vol. 47, no. 3, pp. 299–350, 1998.

[17] D. Mackay, "Gaussian processes - a replacement for supervised neural networks?" *Lecture notes for a tutorial at NIPS*, 1997.

[18] [Online]. Available: http://db.lcs.mit.edu/labdata/labdata.html

[19] R. A. Horn, *Topics in matrix analysis*. New York, NY, USA: Cambridge University Press, 1986.

[20] S. Arora, A. Frieze, and H. Kaplan, "A new rounding procedure for the assignment problem with applications to dense graph arrangement problems," *IEEE Symposium on Foundations of Computer Science*, vol. 00, pp. 21–30, 1996.