Compressed Sensing and Routing in Multi-Hop Networks

Sungwon Lee
EE-Systems, USC
Los Angeles, CA, USA
sungwonl@usc.edu

Sundeep Pattem
EE-systems, USC
Los Angeles, CA, USA
pattem@usc.edu

Maheswaran Sathiamoorthy *
ECE, IIT Kharagpur,
Kharagpur, WB, India
maheswaran@ece.iitkgp.ernet.in

Bhaskar Krishnamachari
EE-systems, USC
Los Angeles, CA, USA
bkrishna@usc.edu

Antonio Ortega†
EE-systems, USC
Los Angeles, CA, USA
ortega@sipi.usc.edu

Abstract

We study how compressed sensing can be combined with routing design for energy efficient data gathering in sensor networks. We first obtain some bounds on the performance of compressed sensing with simple routing schemes. We then formulate a new problem relating routing paths to data projections, and present a centralized, greedy algorithm for obtaining low coherence projections while simultaneously reducing reconstruction error and communication cost. Simulation results show that the effectiveness of standard compressed sensing techniques is limited when routing costs are considered. A naive, randomized downsampling is shown to outperform the standard techniques in terms of achievable SNR for a given energy budget. While our algorithm has better performance than standard compression sensing, it does not improve significantly over downsampling in terms of the cost. Thus we believe that further investigation is needed to determine if novel routing strategies exist that i) provide a sufficient spatial coverage to enable good coherence properties for compressed sensing along those routes, and ii) have transport costs that do not deviate significantly from those of efficient routing techniques.

*Intern at ANRG, USC, Summer 2007.

†The work described here is supported in part by NSF through grants CNS-0347621, CNS-0627028, CCF-0430061, CNS-0325875, and by NASA through an AIST grant. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or NASA.

1 Introduction

Joint routing and compression has been studied for efficient data gathering of correlated sensor network data. Most of the early works were theoretical in nature and while providing important insights, ignored the practical details of how compression is to be achieved [3] [6] [10]. More recently, it has been shown how practical compression schemes such as distributed wavelets can be adapted to combine efficiently with routing [2] [7]. The key issue is to understand the interactions between data structure and routing structure.

Results in compressed sensing [4] [1] provide a radically different view of the structure of data and suggest new approaches to earlier problems. The basic result is that given that an n-dimensional signal is k-sparse in a certain basis, only $O(klogn)$ random projections are required for near-perfect reconstruction. The projections can be obtained using a large class of matrices, including $\pm$ 1 Bernoulli or Gaussian matrices. Subsequent work has extended the results to refinable approximation for compressible signals. Wang *et.al.* [11] showed that the remarkable results of compressed sensing could also be obtained using sparse random projections. The key result they prove is that sparse random projections of vectors preserve the inner product in expectation. They showed that this has important implications for using compressed sensing in a network scenario by reducing the amount of data that needs to be transported to obtain the required projections. However, routing issues in a multi-hop scenario were not considered.

Our work is the first to study how compressed sensing can be combined with routing design for energy efficient data gathering in sensor networks. We present asymptotic analysis of compressed sensing performance with simple routing schemes and an example in which compressed sensing is clearly *not* an effective solution. We then formulate a new problem relating routing paths and data projections and present a centralized, greedy algorithm for obtaining projections with low coherence. The algorithm is motivated by techniques proposed by Elad [5] to iteratively alter the measurement matrix to reduce its coherence with the basis matrix. In our formulation, each row of the projection matrix represents a path in the network and in view of routing costs, many operations in Elad's method are ruled out. Our algorithm is designed for simultaneously reducing reconstruction error and communication cost. We present simulation results comparing the performance of this algorithm and other schemes for routing and compression for data with varying level of compressibility in two different orthogonal basis functions.

Most importantly, our results show that the effectiveness of compressed sensing is limited when routing costs are considered, at least for the two basis functions considered. A naive, randomized downsampling (obtaining data from a subset of nodes) is shown to outperform compressed sensing based techniques in terms of achievable SNR at a given energy budget. While our algorithm has better performance than standard compression sensing, surprisingly, it cannot improve significantly over downsampling. Comparison with a wavelet based routing and compression scheme shows that downsampling, coupled with orthogonal matching pursuit for reconstruction, has better cost performance at low and medium SNR values.

2. Compressed Sensing with Simple Routing Schemes

2.1. Compressed Sensing Basics

Compressed Sensing(CS) builds on the on the surprising revelation that a signal(x) with a length of n having a sparse representation in one basis can be recovered from a small number of projections onto a second basis that is incoherent with the first.

If a signal is sparse in a given basis named sparsity inducing basis,

$$x = \Psi a, |a|_1 = k,$$

then we can reconstruct the original signal with $O(klogn)$ measurements by solving a convex optimization problem,

$$y = \Phi x = \Phi \Psi x = Hx,$$

where H is known as the holographic basis. Reconstruction is possible using H along with the orthogonal matching pursuit algorithm. Measurements are projections of the data onto the measurement vectors,

$$y_i =< \phi_i, x >, \ \phi_i \text{ is a } i^{th} \text{ row of } \Phi.$$

Interestingly, independent and identically distributed (i.i.d.) Gaussian or Bernoulli/Rademacher (random ± 1) vectors provide a useful universal measurement basis that is incoherent with any given Ψ with high probability; the corresponding measurements will capture the relevant information of a sparse signal, regardless of its structure.

CS work can be extended to compressible signals. We say that the data is compressible if the magnitude of its transform coefficients decay like a power law. That is, the i^{th} largest transform coefficient satisfies

$$|\theta_i| \leq R i^{-1/p} \forall i,$$

where

$$|\theta_1| \geq |\theta_2| \geq ... \geq |\theta_n|, \ R \text{ is a constant, and } 0 \leq p \leq 1.$$

Note that p controls the compressibility of the transform coefficients. Wang *et al.* [11] showed that the remarkable results of compressed sensing could also be obtained using sparse random projections(SRPs). The measurement matrix for SRPs is defined as:

$$\Phi_{ij} = \begin{cases} +1 & \text{if } p = \frac{1}{2s} \\ -1 & \text{if } p = \frac{1}{2s} \\ 0 & \text{otherwise} , \end{cases}$$

where s is a parameter that determines the sparseness of the projections and as a result, the number of projections required for good reconstruction. It was shown that with $s = \frac{n}{logn}$ and $O(klogn)$ projections, reconstruction quality is as good as with obtaining the largest k transform coefficients.

2.2. Asymptotic analysis with Simple Routing

Consider a network of n sensor nodes with diameter d hops. The average distance of nodes from the sink is also $O\big(d\big)$ hops. If every node sends its raw sensor measurement to the sink (independently) via the shortest path tree, then the average bit-hop cost per reading for the network is

$$Cost_{raw-SPT} = O(nd).$$

Now consider compressed sensing with dense random projections (DRP). For each projection, every node transmits once and the bit-hop cost is minimized when routing is along a spanning tree . Nodes route data to the sink along this tree. Each node adds its own reading multiplied by ± 1 to the value received from all its children in the tree and sends this new value to its parent. The sink can add values received from each of its children to obtain one complete projection. Since each node in the tree transmits exactly once, the cost per projection is n. Assuming that the projection matrix is known to sink and nodes (each node only needs its column vector) in advance, the cost for obtaining $O(klogn)$ projections using DRPs is

$$Cost_{CS-DRP} = O(n.klogn) = O(knlogn).$$

For obtaining sparse random projections, each node decides to send with probability $\frac{1}{s} = \frac{logn}{n}$ and the measurement is routed along the shortest path. The sink generates the row of the measurement matrix by placing ± 1 at positions for nodes from which data was received and 0 for all others. Since node choice is random, the average path length remains $O(d)$ and the cost using $O(klogn)$ SRPs is

$$\begin{aligned} Cost_{CS-SRP} &= O(d.logn.klogn) \\ &= O(k.d.log^2 n). \end{aligned} \tag{1}$$

This is a bound on the cost that any new scheme based on CS must better.

Observation: In a multi-hop scenario, shortest path routing is efficient for compressed sensing via sparse random projections.

Reasoning: For a single projection, each node decides to send its measurement to the sink with probability $\frac{1}{s} = \frac{logn}{n}$. Reducing the number of bits transported is possible only if data coming form different nodes belongs to the same projection. Since the distribution of the $O(logn)$ (in expectation) nodes that choose to send measurements is random, the chances for coordination in the routing are minimal, and decrease with increasing n. The cost of deviating from the shortest path outweighs the possible gains from reduced bits.

We note that, in general, the optimal routing tree for the node positions in a projection can be expected to be a Minimum Steiner Tree (MST). The Greedy Incremental Tree (GIT) is known to be a good approximation to the MST. By comparing the energy cost for obtaining sparse random projections with the GIT and shortest path, it was verified that the difference is not significant for the network sizes and parameters considered in our simulations.

3. Routing Design for Efficient Compressed Sensing

The bound obtained for SRPs with SPT routing is $O(kdlog^2 n)$. For networks with hundreds or a few thousands of nodes, values of $k, logn$ and d might not be small enough (relative to n) for compressed sensing to provide significant gains compared to raw data transport, which has cost $O(dn)$. Motivated by the work on sparse random projections, we consider route design for computing projections enroute to the sink and obtain good field reconstruction at a cost lower than that using SRPs.

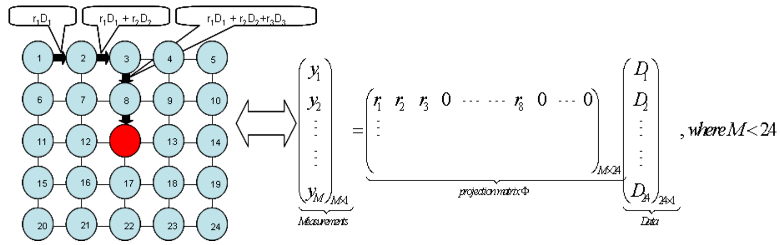


Figure 1. Illustration of the relationship between the routing paths and projections

3.1. Design of measurement matrix

Each row of the measurement matrix represents a path. The projection is computed hop by hop on the way to the sink. This is illustrated in Figure [**?**]. If every route contains only a small number of nodes, the projections will be sparse. The more these routes deviate from the shortest path, the higher the cost. However, because individual nodes do not transport their data separately but combine it with the received value for the partial projection, this leads to energy savings. The key issue is whether a small number of such routes are sufficient for good reconstruction.

The main challenge is that while projections are via rows, incoherence is a property of the columns.

$$\text{Coherence}\mu = \max_{i \neq j} \langle H_i, H_j \rangle.$$

The incoherence between Φ and Ψ is determined by the columns of Φ. The columns of H can be written as:

$$H_i = \Phi\Psi_i, H_j = \Phi\Psi_j.$$

The inner product

$$\langle H_i, H_j \rangle = \Psi_i^T \Phi^T \Phi \Psi_j.$$

For small coherence value, we require $\Phi^T\Phi \approx I$. Clearly this is a condition on the columns, requiring them to be approximately orthogonal. If this condition is satisfied and since Ψ is a orthogonal basis matrix,

$$\langle H_i, H_j \rangle = \Psi_i^T \Phi^T \Phi \Psi_j \approx \Psi_i^T \Psi_j = \delta(i-j)$$

Currently, we do not have a good solution to the problem of systematically achieving low coherence via structure of the basis Ψ. We now investigate a heuristic method for designing routes to obtain quasi random projections and low coherence.

3.2. Low Coherence Projections for Efficient Routing (LCPR)

We present a centralized algorithm for route design, for which idea is to iteratively build paths (which correspond to projections), making a greedy choice for the next node which minimizes the "intermediate coherence" of the partial measurement matrix ($\Phi_{partial}$) with a given Ψ. If it is known that while the readings change over time, the field remains sparse/compressible in a particular basis, then this computation of the routes needs to be performed only once. There is a one time cost for propagating the routing information.

For each projection, start with a randomly chosen node($Node_{init}$) in the network. Iteratively make a greedy choice for the next node on the path, by identifying another node within communication range(R_c) of the current node that will minimize the coherence with the updated $\Phi_{partial}$. In addition to minimizing coherence (μ), we use some heuristic constraints for evaluating candidates to reduce energy consumption.

- node must be closer to sink than the current node
- "crossing over" the sink is not allowed

For each projection, this terminates when there are no neighbors that can reduce the coherence. On path termination, the projection obtained at the last node is routed to the sink along the shortest path. Pseudo-code for the algorithm is in the Appendix.

Our algorithm is motivated by techniques proposed by Elad [5] to iteratively alter the measurement matrix to reduce its coherence with the basis matrix.

We now present an example which indicates that the structure of the sparsity (or compressibility) inducing basis might be a factor that determines whether or not compressed sensing provides improved performance. Consider the scenario with spatial sparsity - at any time, at most k out of n sensors have non-zero data to send. This might happen in a network with sensors that set off alarms when their measurements exceed a threshold. In this case, the sparsity inducing basis $\Psi = I$, the identity matrix.

Proposition 3.1 *Compressed sensing is ineffective for data gathering in case of spatial sparsity.*

Reasoning: If we use sparse random projections, $Cost_{CS-SRP} = O\big(kdlog^2n\big)$. However, if only nodes that set off alarms route their measurements via shortest path to the sink, the cost is $O\big(kd\big)$.

This example is an early indicator that the effectiveness of compressed sensing might be limited when routing costs are taken into account. In the next section, we systematically study routing costs for compressed sensing and other schemes for compression.

4. Simulation Results

We analyze the performance of the various schemes using MATLAB simulations. The experiment details are described first.

4.1. Experiment Details

4.1.1 Routing and compression schemes

We compare the following schemes:

1. the proposed algorithm (LCPR)
2. Sparse random projections via shortest path, with opportunistic projection computations/compression. As argued in Section. 2, this is close to optimal in terms of routing cost.
3. randomized downsampling (DS), routed via shortest path. For SRPs obtained with large s values, the measurement matrix is close to one that results in a simple, randomized downsampling. Hence we also look at the performance of downsampling.
4. projection augmented downsampling (ADS), routed via shortest path. In this scheme, a random subset of nodes is chosen and their data is routed to sink along the shortest path. Nodes on the path contribute their data to the projection with some probability.
5. 2D wavelet based scheme developed by Shen and Ortega [7], invertible transform, optimal assignment of compression levels, shortest path routing, code provided by authors.

DRP is not considered since the cost of obtaining projections is prohibitive. For the CS based algorithms and downsampling, reconstruction is via Orthogonal Matching Pursuit algorithm [9] provided in the SparseLab package [8].

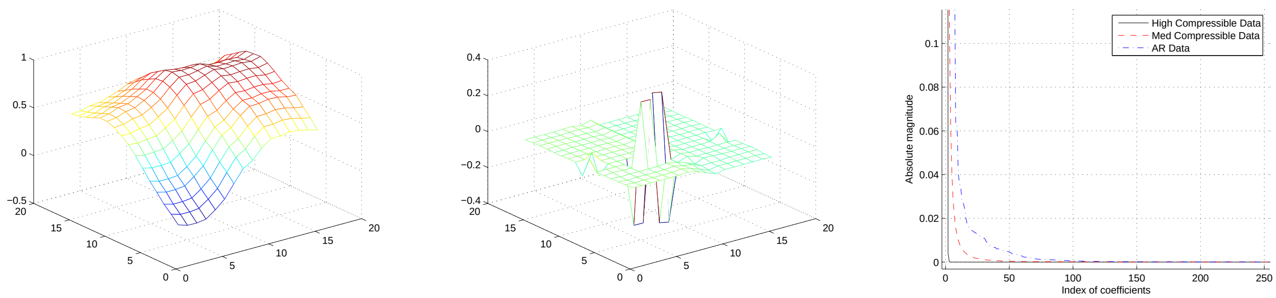
4.1.2 Metrics

The two metrics for performance are as follows:

1. The reconstruction quality is measured using the Signal-to-Noise-Ratio.
2. To study compression gains, we also look at the ratio of algorithm cost to the cost of gathering raw (quantized) data using shortest path routing.

   The energy cost per transmission is obtained as

   1. bit-hop metric



(a) (b) (c)

Figure 2. Sample of data generated using (a) DCT basis (b) Haar basis (c) Decay profile of transform coefficients with high ($p = \frac{1}{8}$), medium ($p = \frac{1}{2}$) and low ($p = \frac{7}{8}$) compressibility

2. the product of the number of bits (b_e) and the edge length (l_e). The cost for a given set of paths is

$$\sum_{p \in P} \sum_{e \in p} b_e . d_e^2 .$$

The first node on a path transmits B bits and we assume that for every new node on the path that contributes a non-zero value, one extra bit is required.

4.1.3 Network topology

The nodes are assumed to be in a square grid deployment. We consider two network size - 256 nodes(16x16) and 1024 nodes(32x32). Nodes can communicate with their 8 adjacent neighbors on the grid, except for those on the corners which have smaller number of neighbors. The sink is located at the center of the square. There is no packet loss.

4.1.4 Data Generation

The data is generated using two different basis functions, 2D Discrete Cosine Transform (DCT) and multi-resolution 2D Haar basis. The magnitude of sorted transform coefficients is given by

$$|\theta_i| = i^{-\frac{1}{p}} .$$

The smaller the value of p, higher the decay rate and hence the higher compressibility [11]. In our experiments we generate data with three different levels of compressibility, $p = \frac{1}{8}, \frac{1}{2}, \frac{7}{8}$. Figure 2 shows samples of data generated with the two bases and the coefficient decay profile for the three values of p.

4.2. Results

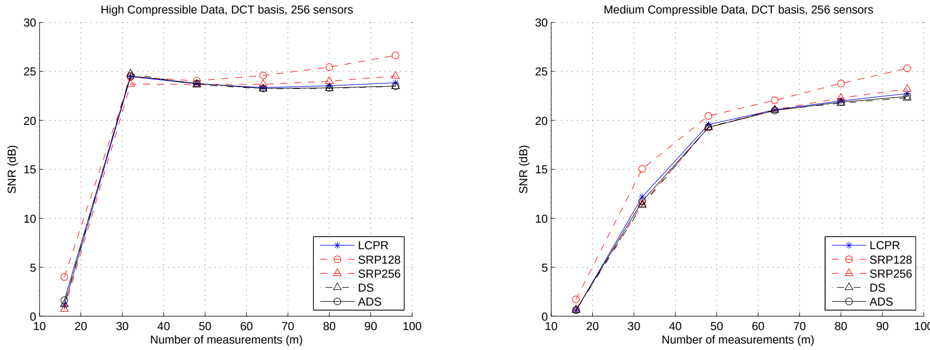
The simulation results presented are averaged over 20 data sets with 10 simulation runs for every data set for each scheme.

Figures 3 and 4 show the SNR performance as a function of number of measurements. For downsampling, we refer to number of nodes chosen for sending data as number of measurements. For DCT basis, SNR performance of SRPs is slightly better than the other schemes. In Figure 4 for Haar basis, the SNR performance of SRPs is better than for downsampling and the algorithm, which have similar performance. The difference is seen to narrow going from high to medium compressibility. These results are in tune with what can be expected from the theory of compressed sensing.

Figures 5, 6 show SNR vs. energy ratio (compared to raw data transport via shortest path) for the schemes with DCT and Haar basis respectively. In all cases, we see the surprising result that downsampling outperforms CS based techniques by achieving higher SNR at same cost. The performance of our algorithm (LCPR) is very close to downsampling. When considered along with the above result for SNR vs number of measurements, this means that the SNR gains of standard compressed sensing techniques are heavily dominated by the routing cost. The naive randomized downsampling combined with a simple routing scheme - shortest path, always has better overall performance.

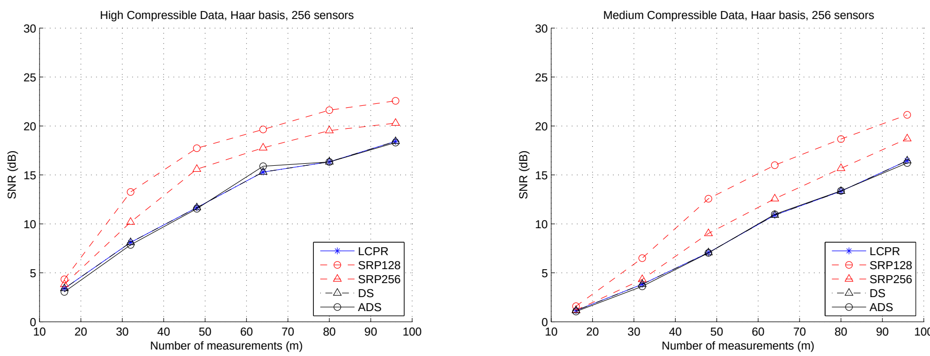
The algorithm achieves lower coherence than downsampling, but this does not seem to be translating to a difference in performance. (Figures 7 (a) and (b). For DCT basis, the coherence with our algorithm and downsampling are mostly lower than with DRPs and SRPs. For Haar basis, the coherence is close to 1 to start with but falls rapidly after a certain number of measurements are available.

Figure 8 is an illustration of the paths chosen by the algorithm for DCT basis with 16 projections. The red stars are starting nodes in each projection, green squares are nodes corresponding to non-zero entries in the projection vector. Circles linked by the paths are relay nodes corresponding to zeros in the projection vector. To reduce clutter, paths with only starting node having non-zero contribution are not



(a) (b)

Figure 3. SNR vs. number of measurements with DCT basis for (a) high and (b) medium compressibility



(a) (b)

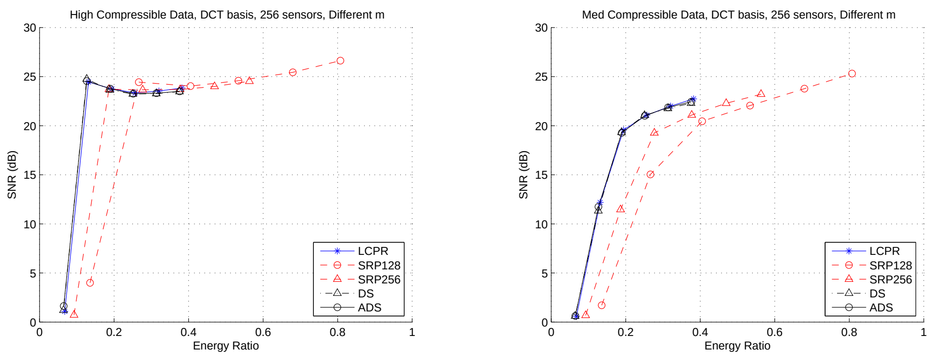
Figure 4. SNR vs. number of measurements with Haar basis for (a) high and (b) medium compressibility

shown. With DCT basis, the algorithm chooses some projections with more than one node and follows paths slightly different from the shortest path, but overall performance and cost is very close to downsampling. With Haar basis, almost all projections chosen by the algorithm consistently have only one node - consequently the measurement matrix is almost a permutation matrix, as for the downsampling case.

Figure 9 shows the relative performance of CS based and 2D wavelet based [7] algorithms. For the low SNR region, our algorithm can provide a higher SNR at the same cost. The 2D wavelet scheme involves an orthogonal transform which preserves the quantization error. With an increasing bit budget, it is possible to obtain all the coefficients more and more accurately and hence, better and better SNR performance. However, with compressed sensing the achievable SNR for compressible data is limited (unless with relatively large number of projections) as mentioned earlier.

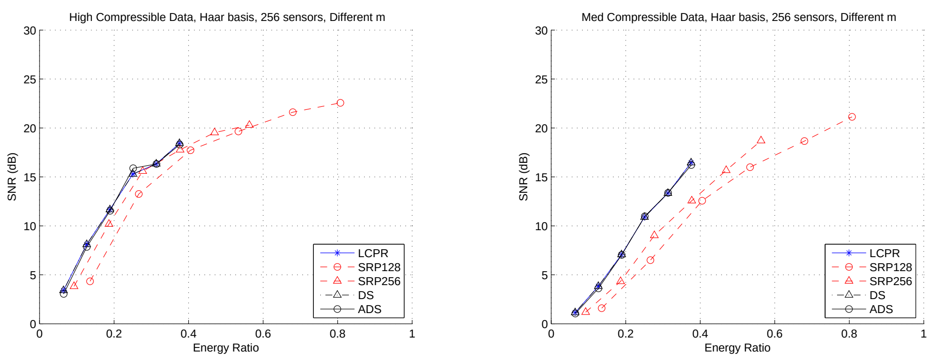
5. Conclusion and Future Work

Standard CS techniques (e.g., DRP, SRP), where the projections have been designed without taking routing costs into consideration, may not be competitive in terms of reconstructed signal representation for given transport costs. Instead, simple techniques such DS and ADS can provide good performance. DS and ADS can be thought of as simplified CS techniques where the transport costs are forced to be minimal (because aggregation is performed only on existing efficient routes). Thus we believe that further investigation is needed to determine if novel routing strategies exist that i) provide a sufficient spatial coverage to enable good coherence properties for compressed sensing along those routes, and ii) have transport costs that do not deviate significantly from those of efficient routing techniques.



(a) (b) (c)

Figure 5. SNR vs. Energy Ratio performance with DCT basis for (a) high and (b) medium compressibility



(a) (b) (c)

Figure 6. SNR vs. Energy Ratio performance with Haar basis for (a) high and (b) medium compressibility

6. Acknowledgements

Thanks to Godwin Shen (USC), for providing code for the 2D wavelet scheme.

References

[1] E. Candes, J. Romberg, and T. Tao. Robust uncertainity principles : exact signal reconstruction from highly incomplete frequency information. In *IEEE Transactions on Information Theory*, pages 489–509. IEEE, Feb. 2006.

[2] A. Ciancio, S. Pattem, A. Ortega, and B. Krishnamachari. Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In *Proceedings of the ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN), Palo Alto, CA, USA*. Springer Verlag, Apr. 2006.

[3] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. On network correlated data gathering. In *Proceedings of the 23rd Conference of the IEEE Communications Society (INFOCOM 2004), Hong Kong*. IEEE Communications Society, Mar. 2004.

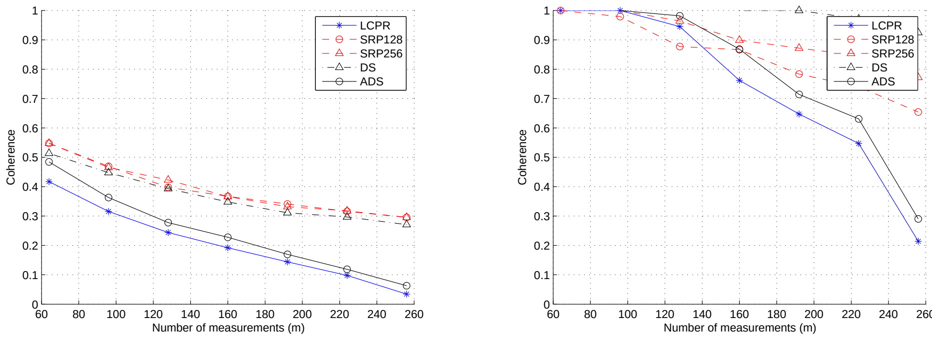
[4] D. L. Donoho. Compressed sensing. In *IEEE Transactions on Information Theory*, pages 1289–1306. IEEE, 2006.

[5] M. Elad. Optimized projections for compressed sensing. In *IEEE Transactions on Signal Processing*.

[6] S. Pattem, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. In *ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN 2004), Palo Alto, CA, USA*, pages 28–35. Springer-Verlag, Apr. 2004.

[7] G. Shen and A. Ortega. Joint routing and 2d transform optimization for irregular sensor network grids using wavelet lifting, submitted ipsn 2008.

[8] Stanford. Sparselab software package. *http://sparselab.stanford.edu/*, 99(7):1–100, January 1999.



(a) (b)

Figure 7. Coherence trend of measurement matrix, with increasing number of projections for (a) DCT basis (b) 2D Haar basis

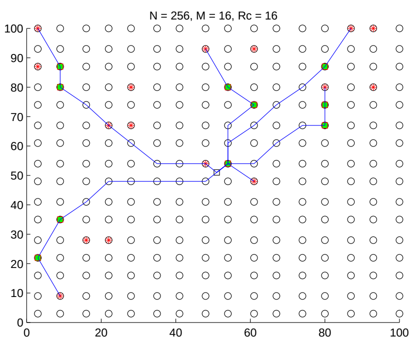


Figure 8. Illustration of routing paths chosen by algorithm for DCT basis

[9] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. In *IEEE Transactions on Information Theory*.

[10] P. von Rickenbach and R. Wattenhofer. Gathering correlated data in sensor networks. In *Proceedings of the DIALM-POMC Joint Workshop on Foundations of Mobile Computing, Philadelphia, PA, USA*, pages 60–66. ACM, Oct. 2004.

[11] W. Wang, M. Garofalakis, and K. Ramachandran. Distributed sparse random projections for refinable approximation. In *Proceedings of the ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN), Palo Alto, CA, USA*, pages 331–339. Springer Verlag, Apr. 2007.

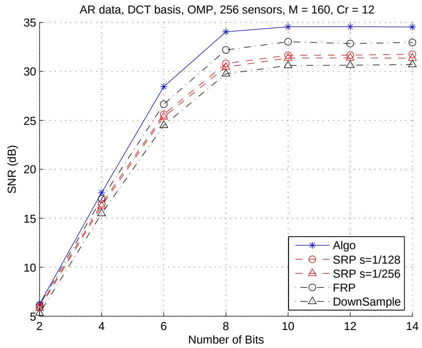
7. Appendix

7.1. Pseudo-code for LCPR algorithm

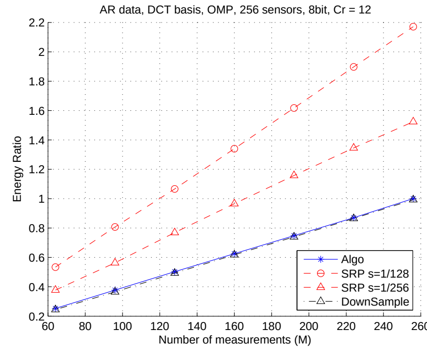
Algorithm 1 LCPR

Require: Ψ, R_c, Nx2 vector Pos_{nodes} consisting of x and y coordinates of N nodes, Pos_{sink} of the sink.
Ensure: Φ, μ.

```
{Find a measurement matrix with low coherence in iterative manner.}
Parent = FindSPT(Pos, R_c)
Determine a set Node_init containing IDs of randomly selected M different nodes.
μ ← 1
for every Node in Node_init do
   Clear Candidates
   FindCandidates(node, R_C, Candidates)
   for every cand in Candidates do
      a=FindNextNode(Candidates, Ψ, Φ_partial)
   end for
end for
if BP.benefit > 0 then
   add BP.node to R and update R accordingly
end if
```



(a)



(b)

Figure 10. Performance comparison for schemes based on compressed sensing for AR data. (a) SNR comparison with varying bit allocation (b) Energy ratio with varying M.

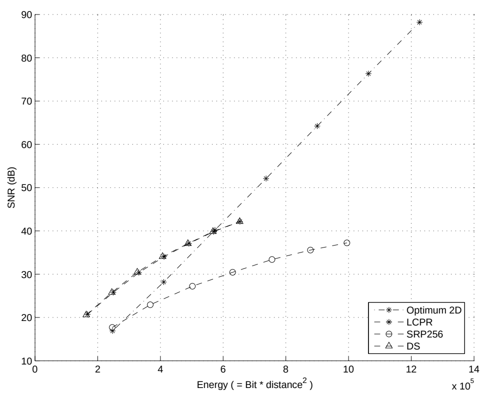


Figure 9. CS vs. 2D wavelet transform with different number of measurements

7.2. Bit allocation, Energy ratio, Large number of nodes

Figure 10(a) shows the performance of the different CS based algorithms with varying bit allocation for data quantization, for a fixed number of projections (M = 160). With increasing number of bits, the SNR obtained first increases steadily up to 8 bits, but using more bits cannot improve it much further. There are two sources of error - quantization and loss of transform coefficients. Increasing number of bits can reduce quantization error close to zero. However, with a relatively small number of projections, it is not possible to get complete information of coefficients for the compressible data considered, even with fine quantization. Hence, there is a limit on the achievable SNR. Given the trend in Figure 10(a), we use 8 bits for quantization in the other experiments presented.

Figure 10(b) is a plot of the energy cost ratio, with respect to raw (quantized) data collection using shortest path, for the different schemes with varying number of projections. For instance, with 160 projections, our algorithm has a cost ratio of 0.6, or 40% savings. As a result of routing costs, SRPs can possibly be useful only for a limited number of projections (while cost ratio is less than 1).