

# Monitoring Embedded Flow Networks using Graph Fourier Transform Enabled Sparse Molecular Relays

Zhuangkun Wei<sup>1</sup>, Alessio Pagani<sup>2</sup>, Bin Li<sup>3</sup> Weisi Guo<sup>1,2,4\*</sup>

**Abstract**—Many embedded networks are difficult to monitor, such as water distribution networks (WDNs). A key challenge is how to use minimum sparse sensors to measure contamination and transmit contamination data to a hub for system analysis. Existing approaches deploy sensors using multi-objective optimization and transmit the data using ground penetrating waves or fixed-line access. Here, for the first time, we introduce a novel molecular communication relay system, which is able to transmit the data report to the hub via the water-flow of WDN itself, and avoids the complex ground penetrating techniques. A water flow data-driven Graph Fourier Transform (GFT) sampling method is designed to inform the invariant orthogonal locations for deploying the molecular relay sensors. Each sensor encodes information via a DNA molecule that enables the common hub to reconstruct the full contamination information. Numerical simulation validates the proposed system, providing a pathway to integrate MC into macro-scale Digital Twin platforms for infrastructure monitoring.

**Index Terms**—molecular communication, network dynamics, water distribution network, sensor deployment, graph Fourier transform

## I. INTRODUCTION

Networked systems are prevalent across all engineered and natural systems, ranging from telecommunications to gas supply, blood vessels, to mycelium fungus networks. In many such cases, the networks are embedded underground or inside a host that is difficult to access. They are often critical to the stakeholders involved and are at risk from a variety of perturbations. Here, we focus on underground water distribution networks (WDNs). Due to their enormous extent and difficulty in access, cascade contamination flow is difficult to track and quantify, which may affect thousands of households [1]. As such, monitoring WDN contamination spread via sensors is of importance for warning and protection.

There are two main challenges to joint sensing and communications: (1) faithful monitoring networked cascades is time- and scenario-dependent, and (2) data inside pipe networks is hard to extract, especially when deep underground.

Addressing challenge (1), current approaches in finding the minimum data to monitor a WDN is studied in [2], [3], whereby a switching-on/off strategy has been proposed based on the principal component analysis (PCA). However, this

requires a large (or complete) set of sensors to be installed at all  $N$  nodes, which incurs an unacceptable high cost of sensor deployment. Minimal sensor placement via numerical simulations of time-varying contamination dynamics and genetic algorithms [4] yields no statistical guarantees. More recent work in [5] studied the signal sampling and reconstruction when node distribution follows Gauss-Poisson point process, which is less attractive in the face of regular lattice networks that do not follow distributions. Graph theoretic methods consider the topology of the network only (e.g. via the Laplacian operator [6]). However, these approaches do not include the underlying fluid dynamics and assume that the topology dominates. In order to include the fluid dynamic data-structure, a well-researched approach is compressed sensing (CS) [7], which selected sampling nodes ensuring complete reconstruction of signal under Restricted Isometric Property (RIP). However, the number of sensors are still very large (approximately  $(N + K - r) \times r/K \geq r$  for a data matrix  $\mathbf{X}$  of  $N$  nodes and  $K$  time-steps with  $\text{rank}(\mathbf{X}) = r$  [8], [9]). In recent research [9], a novel Graph Fourier Transform (GFT) operator matrix was designed that can further improve the number of sampling sensors to  $r$  was designed. However, the open challenge of (2) remain, which is how to transport the data at the sensors to a hub for signal reconstruction. Other sensor placement methods for similar scenarios (e.g., the blood vessels [10], [11]) also do not consider the data report challenge of (2).

Addressing challenge (2), current underground wireless communication systems use a variety of ground penetrating techniques ranging from ultra-low frequency, underground wave-guides [12], to magnetic induction [13]. The devices required tend to be bulky and risk either antenna damage or are not embedded enough to enable large-scale continuous data collection. Fixed line access is also attractive, but damages to cables due to high underground pressure forces means an embedded wireless solution is attractive.

Current studies on molecular communications (MC) [14], [15] provide an inspiration of using messenger molecules (e.g., the encoded DNA molecule [16]) for data transmission in WDN. In terms of the molecular communication propagation channel, at the nano-scale (low Peclet number), mass diffusion dominates propagation spread [14], [17], [18]. However, when it comes to the macro-scale WDN (high Peclet number and Reynolds number), viscous and turbulent diffusion along with flow velocity dominates the spread mechanism [9], [19]. As such, the design of molecular communication network in WDN requires the consideration of Navier-Stokes dynamics.

In this work, we design a molecular relay system [20], [21]

<sup>1</sup>University of Warwick, UK. <sup>2</sup>The Alan Turing Institute, UK. <sup>3</sup>Beijing University of Posts and Telecommunications, China. <sup>4</sup>Cranfield University, UK. \*Corresponding Author: weisi.guo@warwick.ac.uk. The authors acknowledge funding from the Lloyd's Register Foundation's Programme for Data-Centric Engineering and the Alan Turing Institute under the EPSRC grant EP/N510129/1, as well as Complexity Twin for Resilient Ecosystems (CoTRE) under the EPSRC grant EP/R041725/1. The authors acknowledge Microsoft Corporation for providing cloud resources on Microsoft Azure.

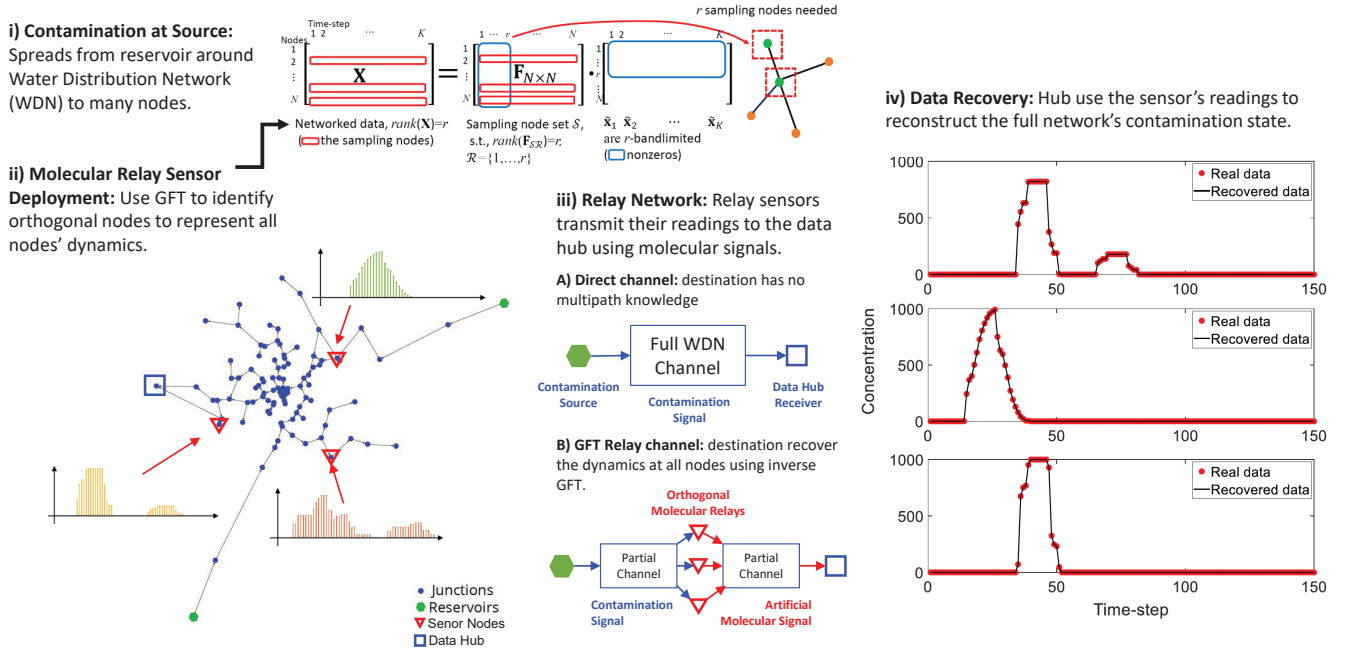


Fig. 1. Molecular relay monitoring system for water distribution network (WDN): (i) contamination spreads via WDN, (ii) GFT operators inform the location of critical sampling points to deploy molecular relays, (iii) a relay network transmit modulated molecular signals via WDN to hub, and (iv) hub reconstructs the full network-wide contamination dynamics using the sensor data.

addressing the aforementioned open challenges by: (1) using DNA-modulated MC in the WDN to transmit the sensor data to the hub, and (2) designing a novel GFT sampling method and optimal sparse sensor deployment for contaminant signal reconstruction. The molecular relay enables the transmissions of sensor reports to the hub via the inherent water-flow of the WDN, and thereby avoids the extra construction of the communication networks and ground penetrating techniques for data transmission. Combined together, we have integrated complex network topology with water flow data-structure, guaranteeing: (a) the existence of paths between all selected nodes to the hub, and (b) error free reconstruction of the full network contamination data. We call this entire system the Molecular Relay Data-Driven GFT (MRDD GFT).

## II. MODEL AND PROBLEM FORMULATION

### A. Network Flow Model

A water-distribution network (WDN) contains the topological structure, and the water-flow dynamics driven by both demand variation and reservoir pressure. In the model, the pipe network is configured by a static graph, denoted as  $\mathcal{G}(\mathcal{V}, \mathbf{W})$ , where  $\mathcal{V} = \{1, \dots, N\}$ ,  $N \in \mathbb{N}^+$  represents the set of indices of total  $N$  nodes, and  $\mathbf{W}$  is the adjacency matrix. Here, the  $(m, n)$ th element  $w_{m,n} \in \{0, 1\}$  of  $\mathbf{W}$  reflects the existence of the directed link from node  $n$  to node  $m$ .

Given the water-flow of the WDN, the contaminant injected by one node will be propagated over the whole network. The aim is to study the minimum number of sensors deployed on the network and how to report such underground samples to a hub without complex penetration techniques, such that the time-varying contaminant data over the whole network can be fully reconstructed. Here, all the sampling nodes and the hub

are nodes (e.g. junctions, reservoirs, pumps) of WDN, which enables the usages of its own pipes for report transmission. We denote  $\mathbf{X}$  of size  $N \times K$  as the contaminant data of total  $N$  nodes and total  $K \in \mathbb{N}^+$  time-steps, and  $\mathcal{S}$  as the sampling node set for sensor deployment. The purpose is to find an appropriate  $\mathcal{S} \subset \mathcal{V}$ , such that

$$\mathbf{X} = \Phi \cdot \mathbf{X}_{\mathcal{S}\mathcal{K}}, \quad (1)$$

where  $\Phi$  is the recovery matrix for signal reconstruction, and  $\mathbf{X}_{\mathcal{S}\mathcal{K}}$  is the samples (submatrix) of  $\mathbf{X}$ , with rows indexed by set  $\mathcal{S}$ , and columns indexed by set  $\mathcal{K} = \{1, \dots, K\}$ .

### B. Background and Previous Work

One way to derive the recovery matrix  $\Phi$  and the sampling node set  $\mathcal{S}$  is leveraged on the graph sampling theory, which can use lesser samples for signal reconstruction compared with other sampling method (e.g., the CS) [22]. It uses the bandlimited property of the networked matrix signal to a Graph Fourier Transform operator matrix, and therefore enables the complete signal reconstruction via samples on  $\mathcal{S}$ .

*Definition 1:* [9] We say data matrix  $\mathbf{X}$  is  $r$ -bandlimited with respect to a GFT operator  $\mathbf{F}^{-1}$ , if the rest  $N - r$  rows of the frequency response  $\tilde{\mathbf{X}} = \mathbf{F}^{-1} \cdot \mathbf{X}$  are all zero vectors. We define the cut-off bandwidth set  $\mathcal{R}$  as the set of subscripts of non-zero rows in frequency response matrix, i.e.,  $\mathcal{R} = \{1, \dots, r\}$ .

*Theorem 1:* [9], [22] For any  $r$ -bandlimited data matrix  $\mathbf{X}$  and its cut-off bandwidth set  $\mathcal{R}$  with respect to a GFT operator  $\mathbf{F}$ , there exists a subset  $\mathcal{S} \subset \mathcal{V}$  such that

$$\mathbf{X} = \mathbf{F}_{\mathcal{V}\mathcal{R}} \cdot (\mathbf{F}_{\mathcal{S}\mathcal{R}}^T \cdot \mathbf{F}_{\mathcal{S}\mathcal{R}})^{-1} \cdot \mathbf{F}_{\mathcal{S}\mathcal{R}}^T \cdot \mathbf{X}_{\mathcal{S}\mathcal{K}}. \quad (2)$$

Such  $\mathcal{S}$  satisfies  $\text{rank}(\mathbf{F}_{\mathcal{S}\mathcal{R}}) = |\mathcal{R}|$ .

Given the graph sampling theory, previous work in [9] formed a data-driven GFT operator  $\mathbf{F}^{-1}$  via the signal space, ensuring the  $r = \text{rank}(\mathbf{X})$ -bandlimitedness of  $\mathbf{X}$ . However, they ignore the practical challenge of how such underground sensors report their data to the hub for signal reconstruction.

### III. MOLECULAR RELAY DATA-DRIVEN GFT

In this section, we elaborate our MRDD GFT sampling system on contaminant propagation over WDN. Here, we adopt the data-driven GFT operator in [9]. The novelty of this work lies in two aspects. First, the design of molecular relay for sensor data transmission avoids the complex ground penetrating techniques (e.g., ultra-low frequency, underground wave-guides [12] and magnetic induction [13]). Second, a novel sampling node selection for sensor deployment is proposed, ensuring the existence of the path between selected nodes and the hub node, such that the reports can be conveyed to the hub for signal reconstruction.

#### A. Molecular Relay

The aim of molecular relay is to transmit the sampled data of sensors to the hub node via the water-flow of WDN itself, other than the construction of extra communication networks using ground penetrating techniques. To do so, we equip each sensor with one specific type of DNA molecule [16], which, for each time-step<sup>1</sup>, encodes the report into the molecular structure, and will be transmitted to the hub via the propagation of water-flow. As such, one prerequisite for critical node selection is to ensure the existence of the paths between each selected node with the hub, so that the hub can receive the molecular report from each sensor.

Due to the different intensities of water flows and water demands in nodes and pipes, the transmitted molecules may be loss, which affects the receiving of the sampled data and will deteriorate the recovery performance (illustrated in Fig. 3). Here, the molecule loss constitutes an erasure channel. Given that the molecular diffusion rate (an order of  $10^{-9}\text{m}^2/\text{s}$  [17], [18]) is much lower than the velocity of the water flow (typically  $1\text{m}^2/\text{s}$  [24]), we neglect the diffusion mechanism but use the fluid parameters to compute the arrival probability. We define the adjacent arrival probability of pair  $(m, n)$  as the receiving probability of one molecule transmitted from node  $n$  to node  $m$  via the pipe  $w_{m,n}$ , i.e.,

$$p_{m,n} = \frac{\Delta_{m,n}}{\sum_{i \in \mathcal{V}} \Delta_{i,n} + D_n} \quad (3)$$

where  $\Delta_{m,n}$  is the average water-flow in pipe  $n$  to  $m$ , and  $D_n$  represents the average water demands of node  $n$ . As such, the arrival probability for one molecule of any pair  $(m, n)$  can be computed by traversing all the path from  $n$  to  $m$ , i.e.,

$$P_{m,n} = 1 - \prod_{\forall l} \left( 1 - \prod_l p_{l_1,n} \cdot p_{l_2,l_1} \cdots p_{m,l_j} \right) \quad (4)$$

where  $l = (n, l_1, \dots, l_j, m)$  is one path from  $n$  to  $m$ .

<sup>1</sup>Sensors are synchronized via the low-rate blind synchronization in [23].

Eq. (4) yields the relationship between the network topology (i.e., the paths) with the arrival probability of each node-pair. For further theoretical sampling node analysis, we assume an extremely large number of transmitted molecules for each sampling node, which thereby converts the arrival probability matrix to a simplified path existence matrix, of which the  $(m, n)$ th element represents whether a path exists from node  $n$  to node  $m$ . Nevertheless, we measure the arrival probability matrix in Fig. 3(a), and show its influence on the proposed scheme in Fig. 3(b).

#### B. Sampling Node Selection

As aforementioned, the non-existence of path from sampling nodes to the hub leads to the report loss, and deteriorates the recovery performance of the existing sampling method in [9].

To address this, we design a novel sampling node selection method, which can be expressed as:

$$(\mathcal{S}, n_{\text{hub}}) \text{ s.t. } \text{rank}(\mathbf{F}_{\mathcal{S}\mathcal{R}}) = |\mathcal{R}|, \quad \left\| (\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{S}} \right\|_{l_0} = |\mathcal{S}|. \quad (5)$$

In Eq. (5),  $n_{\text{hub}}$  denotes the hub node from the nodes of the WDN.  $\mathbf{W}_0 = \mathbf{W} + \mathbf{I}_N$  ( $\mathbf{I}_N$  is an identity matrix of size  $N \times N$ ) is the revision of adjacent matrix  $\mathbf{W}$  by considering self-loop of each node. aiming to analyze the path of each node-pair. The reason of using this self-loop revision  $\mathbf{W}_0$  is to analyze the existence of paths for each node-pair, as the  $L$ -exponent of  $\mathbf{W}_0$ , i.e.,  $\mathbf{W}_0^L$ , can store the existence of any  $\leq L$ -length paths, as opposed to  $\mathbf{W}^L$  stores only the existence of  $L$ -length paths. As such, given a large  $L$ , the  $(m, n)$ th element of  $\mathbf{W}_0^L$  represents the status (i.e., existence or not) of the path from node  $n$  to node  $m$ .  $(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{S}}$  is a sub-matrix of  $(\mathbf{W}_0)^L$  with  $n_{\text{hub}}$  row and columns in  $\mathcal{S}$ , representing the status of the paths from nodes  $\mathcal{S}$  to the node  $n_{\text{hub}}$ . The role of  $l_0$ -norm of  $(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{V}}$  is to count the number of paths from other nodes in  $\mathcal{S}$  to the hub node.  $\|(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{S}}\|_{l_0} = |\mathcal{S}|$  therefore ensures the existence of paths from all sampling nodes in  $\mathcal{S}$  to the hub  $n_{\text{hub}}$ . As such, Eq. (5) is to search the pair of critical nodes set  $\mathcal{S} \subset \mathcal{V}$ , and the hub node  $n_{\text{hub}} \in \mathcal{V}$  that ensures the successful report and the signal reconstruction.

One intuitive way to solve Eq. (5) is given in the following.

(i) We find all potential  $n_{\text{hub}}$  by examining whether the  $l_0$ -norm of  $n_{\text{hub}}$ th row of  $\mathbf{W}_0^L$  is larger than  $|\mathcal{R}|$  i.e.,  $\|(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{V}}\|_{l_0} \geq |\mathcal{R}|$ . This is because for the rows with smaller  $l_0$ -norm,  $\|(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{S}}\|_{l_0} \leq \|(\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{V}}\|_{l_0} < |\mathcal{R}|$  with  $|\mathcal{R}| \leq |\mathcal{S}|$  deduced by  $\text{rank}(\mathbf{F}_{\mathcal{S}\mathcal{R}}) = |\mathcal{R}|$ .

(ii) For each potential  $n_{\text{hub}}$ , we combine the network topology with the GFT operator, and derive a topology-data combined operator as:

$$\mathbf{F}^{(n_{\text{hub}})} = \text{diag}(\mathbf{1}((\mathbf{W}_0^L)_{\{n_{\text{hub}}\}\mathcal{V}})) \cdot \mathbf{F}, \quad (6)$$

where  $\text{diag}(\cdot)$  is to construct a diagonal matrix, and  $\mathbf{1}(\cdot)$  is to map non-zeros of a vector into 1, and zeros into 0. As such, the rows in such topology-data combined operator  $\mathbf{F}^{(n_{\text{hub}})}$  are zeros if their corresponding nodes do not have paths to  $n_{\text{hub}}$ . Hence, using  $\mathbf{F}^{(n_{\text{hub}})}$  for the further rank analysis and sampling node selection guarantees the conditions in Eq. (5).

(iii) We select the node as hub whose matrix  $\mathbf{F}_{\mathcal{V}\mathcal{R}}^{(n_{\text{hub}})}$  has minimum condition number, i.e.,

$$n_{\text{hub}} = \underset{n \in \mathcal{V}}{\operatorname{argmin}} \operatorname{cond} \left( \mathbf{F}_{\mathcal{V}\mathcal{R}}^{(n)} \right) \quad (7)$$

where  $\operatorname{cond}(\cdot)$  represents the condition number.

(iv) For the selected  $n_{\text{hub}}$ , we aim to find  $\mathcal{S} \subset \mathcal{V}$  such that  $\operatorname{rank}(\mathbf{F}_{\mathcal{S}\mathcal{R}}^{(n_{\text{hub}})}) = |\mathcal{R}|$  in order to ensure the invertibility of  $\mathbf{F}_{\mathcal{S}\mathcal{R}}^{(n_{\text{hub}})}$ . This is equivalent to minimize the condition number or maximize the minimum singular value of  $\mathbf{F}_{\mathcal{S}\mathcal{R}}^{(n_{\text{hub}})}$ , i.e.,

$$\mathcal{S} = \underset{\mathcal{S} \subset \mathcal{V}}{\operatorname{argmax}} \sigma_{\min} \left( \mathbf{F}_{\mathcal{S}\mathcal{R}}^{(n_{\text{hub}})} \right), \quad (8)$$

where  $\sigma_{\min}(\cdot)$  denotes the smallest singular. Given that Eq. (8) is a NP-hard problem, we hereby use a *greedy* algorithm, in which one can identify and add the  $n$ th row, i.e.,  $\mathcal{S} \leftarrow \mathcal{S} \cup \{n\}$ , such that  $n = \operatorname{argmax}_n \sigma_{\min}(\mathbf{F}_{(\mathcal{S}+\{n\})\mathcal{R}}^{(n_{\text{hub}})})$ .

### C. Signal Reconstruction

The signal reconstruction process can be pursued by taking  $\mathbf{F}_{\mathcal{S}\mathcal{R}}^{(n_{\text{hub}})}$  and  $\mathcal{S}$  into Theorem 1. Here, in order to further address the report loss in  $\mathcal{S}$ , we use the last arrival report as a replacement. To be specific, suppose the report from node  $s \in \mathcal{S}$  is lost at  $k$  time-step, then we use its last arrival report to replace current lost report, i.e.,  $\mathbf{X}_{s,k} = \mathbf{X}_{s,k'}$  where  $k' < k$  is the last time-step when report from node  $s$  is received.

## IV. RESULTS

In this section, we evaluate the reconstruction performance of the proposed MRDD-GFT scheme, in terms of (i) the molecular communication connectivity, and (ii) the reconstruction accuracy. The simulation configuration is given as follows. For the WDN topology, we use one of the typical topology provided from industry standard EPANET2 [24], which has  $N = 102$  nodes (see Fig. 1), consisting of 100 junctions and 2 reservoirs. Then, we use industry standard EPANET2 [24] to simulate the water and the contaminant. We assign 100 random water-demands, and use Computational Fluid Dynamics (CFD) solutions to approximate real flow in pipes and duct channels. The contaminant is simulated by an initialized injection at one node and spread via the water-flow propagation. For molecular relay, we compute the arrival probability given the data of water-demands and water-flows in pipes using Eqs. (3)-(4). Then, we transmit  $10^4$  molecules for each report from each sampling node to the hub.

### A. Molecular Communication Connectivity

We firstly test the existence of hub for the proposed MRDD GFT, as the existence of the hub for all selected nodes ensures the successful collection of the reports conveyed by molecular relays for signal reconstruction. In Fig. 2, the ratio of sampling nodes linking to the hub is provided (i.e.,  $|\mathcal{S}_{\text{hub}}|/|\mathcal{S}|$ ) with respect to the size of the sampling node set  $|\mathcal{S}|$ . It is observed that for the proposed MRDD GFT method, such ratio is 1 for  $|\mathcal{S}| \leq |\mathcal{S}_{\text{min}}| = 41$  combined with Fig. 2(a), as opposed to the ratio of existing proposed data-driven GFT method in [9] that reduces to 0.5 for  $|\mathcal{S}| \leq |\mathcal{S}_{\text{min}}| = 41$ . This indicates that

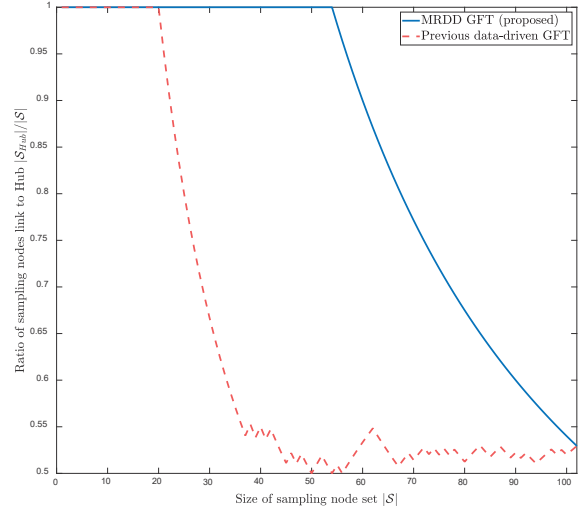


Fig. 2. Comparison of whether nodes in sampling node set can have a hub between the proposed MRDD GFT and existing data-driven GFT [9]. The new MRDD GFT can enable a much larger set of sampling nodes to be connected (via molecular communication) to the hub than the existing GFT design [9].

the existing data-driven GFT method in [9] cannot guarantee the existence of the hub for all sampling nodes, thereby inevitably leading to the loss of reports for data reconstruction. By contrast, the proposed MRDD GFT method is capable of maintaining the existence of the hub for the successful collection of the reported data from the requested sampling nodes (i.e.,  $|\mathcal{S}| = r = 41$  for this case), and therefore data reconstruction can be ensured.

The guarantee of the existence of the hub for the proposed MRDD GFT method is attributed to the combination of the topology information of the network with the data-driven GFT operator in Eq. (6). As such, the node selection in Eq. (8) ensures (i) the existence of the hub and (ii) the complete data reconstruction.

### B. Reconstruction Accuracy

The reconstruction accuracy is measured by the normalized root mean square error (NRMSE) of the reconstructed data in the face of erasure molecular channel. We provide the molecule arrival probability of all pairs in Fig. 3(a), and the reconstruction accuracy of the proposed MRDD GFT sampling method is given in Fig. 3(b).

In Fig. 3(b), it is firstly shown that the MRDD GFT method outperforms the existing data-driven GFT method in [9] in the absence of the molecule (report) loss. This is because the proposed MRDD-GFT method is able to ensure the existence of the paths from all the sampling nodes to the hub, which constitutes the prerequisite that the hub can receive all the required samples for signal reconstruction. Then, we can observe that in the presence of the molecule erasure, the NRMSE of the MRDD GFT is still lower than that of the existing data-driven GFT in [9]. Here, such advantage comes from two aspects. First, this is attributed to the connectivity between sampling nodes and the hub that guarantees the non-zero arrival probabilities of report molecules. Second, when the report molecules lose, the utilization of previous arrival

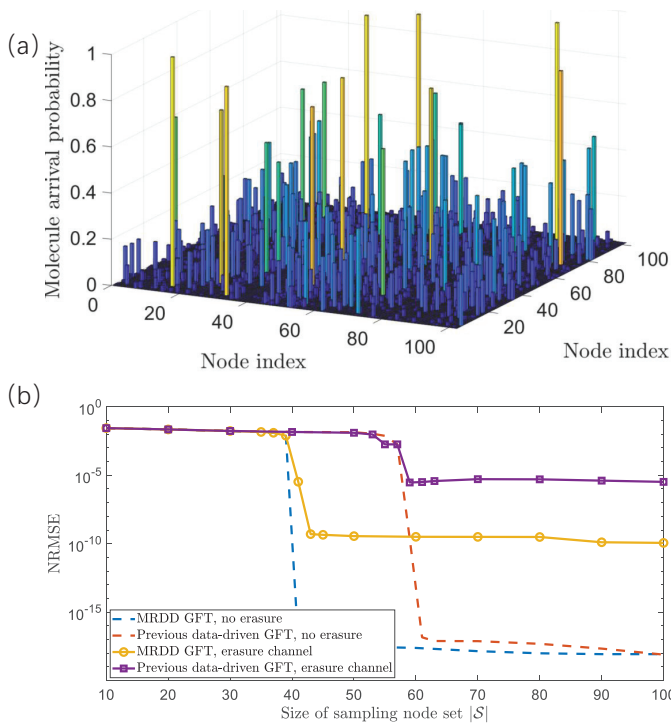


Fig. 3. Comparison of reconstruction accuracy among the proposed MRDD GFT and previous data-driven GFT in [9]. (a) gives the one molecule arrival probabilities for any pair nodes. (b) provides the normalized RMSE.

reports in Section III. C contributes to the lower NRMSE of the proposed MRDD-GFT method.

## V. CONCLUSIONS

Monitoring the contaminant propagation over Water distribution networks (WDN) is challenging due to the high dimensional dynamics, the large-scale network involved, and the underground embedding. This raises challenges on the optimal (minimum) sensor placement, and their transmission of data to a hub. In this work, we introduce a novel molecular relay based data-driven Graph Fourier Transform (MRDD GFT), which allows us to fully reconstruct the contaminant data by using a subset of data sampling at critical nodes. This approach uses the molecular propagation via the water-flow to transmit the sensor reports to the hub, which addresses the challenges of data access for underground sensors. We use standardized fluid dynamic simulations across a complex network topology to validate the results, providing a pathway to integrate MC into macro-scale Digital Twin platforms for infrastructure monitoring. Future work will focus on how to improve the trade-off between minimum sensor-side processing, edge and cloud computing, and big data fusion for intelligent water distribution networks and Digital Twin modelling.

## REFERENCES

- [1] V. Pye and R. Patrick, "Ground water contamination in the United States," *Science*, vol. 221, no. 4612, pp. 713–718, Aug. 1983.
- [2] R. Du, L. Gkatzikis, C. Fischione, and M. Xiao, "Energy efficient sensor activation for water distribution networks based on compressive sensing," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2997–3010, Dec. 2015.

- [3] Z. Wei, A. Pagani, and W. Guo, "Monitoring networked infrastructure with minimum data via sequential graph fourier transforms," in *IEEE International Smart Cities Conference*, Oct. 2019.
- [4] N.-B. Chang, N. Prapinpongsonane, and A. Ernest, "Optimal sensor deployment in a large-scale complex drinking water network: Comparisons between a rule-based decision support system and optimization models," *Comput. Chem. Eng.*, vol. 43, pp. 191–199, Aug. 2012.
- [5] F. Zabini, G. Pasolini, and A. Conti, "On random sampling with nodes attraction: The case of Gauss-Poisson process," in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 2278–2282.
- [6] C. Giudicianni, A. Nardo, M. Natale, R. Greco, G. Santonastaso, and A. Scala, "Topological taxonomy of water distribution systems," *Water*, vol. 10, no. 4, p. 444, Apr. 2018.
- [7] S. Kartakis, G. Tzagkarakis, and J. A. McCann, "Adaptive compressive sensing in smart water networks," in *2nd International Electronic Conference on Sensors and Applications*, 2015.
- [8] E. J. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.
- [9] Z. Wei, A. Pagani, G. Fu, I. Guymer, W. Chen, J. A. McCann, and W. Guo, "Optimal sampling of water distribution network dynamics using graph fourier transform," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2019.
- [10] T. Matsunaga, Y. Haga, S. Osaki, Y. Shimizu, S. Tupin, M. Ohta, M. Shojima, H. Yoshida, and N. Tsuruoka, "Multipoint pressure measurement in blood vessel model for evaluation of intravascular treatment of cerebral aneurysm using fiber-optic pressure sensors," in *IEEE International Conference on Cyborg and Bionic Systems*, Oct. 2017, pp. 136–139.
- [11] B. Ibrahim, D. Mrugala, and R. Jafari, "Effects of bio-impedance sensor placement relative to the arterial sites for capturing hemodynamic parameters," in *IEEE International Conference in Medicine and Biology Society*, Jul. 2019, pp. 6569–6573.
- [12] S. Qiu, C. Mias, W. Guo, and X. Geng, "Hs2 railway embankment monitoring: effect of soil condition on underground signals," *SN Appl. Sci.*, vol. 1, no. 6, p. 537, May 2019.
- [13] Z. Sun and I. F. Akyildiz, "Magnetic induction communications for wireless underground sensor networks," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 7, pp. 2426–2435, Jul. 2010.
- [14] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Trans. Mol., Biol. Multi-Scale Commun.*, vol. 2, no. 2, pp. 120–142, Dec. 2016.
- [15] F. Zabini, "Spatially distributed molecular communications: An asynchronous stochastic model," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1326–1329, Jul. 2018.
- [16] B. A. Bilgin, E. Dinc, and O. B. Akan, "Dna-based molecular communications," *IEEE Access*, vol. 6, pp. 73 119–73 129, 2018.
- [17] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2532–2547, Jun. 2011.
- [18] Z. Wei, W. Guo, B. Li, J. Charmet, and C. Zhao, "High-dimensional metric combining for non-coherent molecular signal detection," *IEEE Transactions on Communications*, pp. 1–1, 2019.
- [19] D. T. McGuinness, S. Giannoukos, A. Marshall, and S. Taylor, "Experimental results on the open-air transmission of macro-molecular communication using membrane inlet mass spectrometry," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2567–2570, Dec. 2018.
- [20] N. Tavakkoli, P. Azmi, and N. Mokari, "Performance evaluation and optimal detection of relay-assisted diffusion-based molecular communication with drift," *IEEE Transactions on Nanobioscience*, vol. 16, no. 1, pp. 34–42, Jan. 2017.
- [21] S. Yuan, J. Wang, and M. Peng, "Performance analysis of reversible binding receptor based decode-and-forward relay in molecular communication systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 880–883, Oct. 2018.
- [22] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.
- [23] Z. Luo, L. Lin, W. Guo, S. Wang, F. Liu, and H. Yan, "One symbol blind synchronization in simo molecular communication systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 530–533, Aug. 2018.
- [24] L. Rossman, "Epanet 2 users manual," *U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-00/057*, 2000.

# Monitoring embedded flow networks using graph Fourier transform enabled sparse molecular relays

Wei, Zhuangkun

2020-03-06

Attribution-NonCommercial 4.0 International

---

Wei Z, Pagani A, Li B, Guo W. (2020) Monitoring embedded flow networks using graph Fourier transform enabled sparse molecular relays. *IEEE Communications Letters*, Volume 24, Issue 5, May 2020, pp. 986-990

<https://doi.org/10.1109/LCOMM.2020.2978835>

*Downloaded from CERES Research Repository, Cranfield University*