Graph Signal Compression by Joint Quantization and Sampling

Pei Li, Nir Shlezinger, *Member, IEEE*, Haiyang Zhang, *Member, IEEE*, Baoyun Wang, *Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Graph signals arise in various applications, ranging from sensor networks to social media data. The high-dimensional nature of these signals implies that they often need to be compressed in order to be stored and transmitted. The common framework for graph signal compression is based on sampling, resulting in a set of continuous-amplitude samples, which in turn have to be quantized into a finite bit representation. In this work, we study the joint design of graph signal sampling along with quantization, for graph signal compression. We focus on bandlimited graph signals, and show that the compression problem can be represented as a task-based quantization setup, in which the task is to recover the spectrum of the signal. Based on this equivalence, we propose a joint design of the sampling and recovery mechanisms for a fixed quantization mapping, and present an iterative algorithm for dividing the available bit budget among the discretized samples. Furthermore, we show how the proposed approach can be realized using graph filters combining elements corresponding the neighbouring nodes of the graph, thus facilitating distributed implementation at reduced complexity. Our numerical evaluations on both synthetic and real world data shows that the joint sampling and quantization method yields a compact finite bit representation of high-dimensional graph signals, which allows reconstruction of the original signal with accuracy within a small gap of that achievable with infinite resolution quantizers.**

***Index Terms*—Graph signal compression, task-based quantization, graph filter, sampling, bit allocation.**

I. Introduction

RAPID development of information and communication technology, has lead to a growing need to process and store high-dimensional signals [2]. In many families of signals, such as those representing communication networks, social media data, and sensors deployment, the interactions between the elements of the signal obey some graphical structure. Graph signal processing (GSP) has emerged as a promising technique to deal with such complex signals [3], [4].

The high-dimensional nature of graph signals gives rise to the need of compressing them [5]. A leading strategy to compress graph signals is based on sampling their elements [6], [7], [8]. Sampling techniques typically build upon the frequency analysis of graph signals, which is a fundamental tool in GSP, utilized also for processing such as denoising [9], [10] and interpolation [11]. Generally speaking, the graph signal values associated with the two end vertices of edges with large weights in the graph tend to be similar [12]. This property, often observed in practice, leads to spectral sparsity, and the resulting signals are referred to as *bandlimited* [13]. Basic graph sampling theory focuses on bandlimited graph signals and relates the spectral support to the number of samples required for representing the signal such that it can be reconstructed from noiseless samples [6]. Nonetheless, in the presence of noise, it is often challenging to determine how to select which nodes to sample. In general, sampling set selection is an NP-hard issue, which is often tackled using greedy approaches [14]. Recently, sampling theorems for non-bandlimited graph signals, exploiting sparsity in domains other than the spectral domain, were proposed in [7], [8].

A key characteristic of graph signal compression by sampling stems from the fact that it represents the signal by a set of continuous-amplitude samples, where the number of samples is treated as the compression dimension. However, when storing and transmitting graph signals, the level of compression is measured in the number of bits, rather than continuous-amplitude samples, required for representing the signal. This implies that graph signal compression should involve not only sampling, but also *quantization* [15]. To date, existing works on quantization in GSP, e.g., [16], [17], [18], [19], [20], focus on applying GSP methods to graph signals with quantized elements [16], [17], [18], [19] or quantizing sampled graph signals [20], rather than designing joint sampling and quantization methods for compressing such signals. This motivates the design of graph signal compression schemes which combine both sampling as well as quantization designed in a joint manner as papers on ADC compression.

In this work, we propose a compression method for bandlimited graph signals which maps a high-dimensional signal into a finite-bit representation via sampling and quantization. Our compression method is inspired by the recently proposed task-based quantization framework [21], [22], [23], [24],

Manuscript received 24 October 2021; revised 14 June 2022 and 18 August 2022; accepted 27 August 2022. Date of publication 9 September 2022; date of current version 26 September 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David I. Shuman. This work was supported in part by the National Natural Science Foundation of China under Grant 61971238, in part by the European Research Council (ERC) through the European Union's Horizon 2020 research and innovation programme under Grant 101000967, and in part by Israel Science Foundation under Grant 0100101 [DOI: 10.1109/ICASSP39728.2021.9414657]. *(Corresponding author: Baoyun Wang.)*

Pei Li, Haiyang Zhang, and Baoyun Wang are with the School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: 2017010211@njupt.edu.cn; 20220142@njupt.edu.cn; bywang@njupt.edu.cn).

Nir Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (e-mail: nirshl@bgu.ac.il).

Yonina C. Eldar is with the Faculty of Math and CS, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/TSP.2022.3205474

[25], [26]. Task-based quantization studies the acquisition of multivariate continuous-amplitude signals in order to recover some underlying information vector, while operating under an overall bit budget. This is achieved by processing the signal using a dimensionality-reducing combining mapping, carried out in the analog domain, followed by scalar quantization and digital processing, all jointly designed to recover the task vector [25]. Our ability to treat graph signal compression as task-based quantization stems from the observation that graph signal sampling can be viewed as an analog combining operation, while the task in reconstructing bandlimited graph signals is the recovery of their spectrum. Unlike the original task-based quantization formulation, which focused on the design of analog-to-digital convertors (ADCs), and thus considered identical quantization mappings applied to each sampled element, here we consider a compression problem rather than signal acquisition, and thus does not enforce this restriction.

Our proposed sampling and quantization design employs a graph filter prior to the quantization operation. We consider two types of such graph filters. The first is an unconstrained graph filter, which is allowed to carry out arbitrary linear operations on the graph signals during the sampling procedure. For the unconstrained setup, we begin by studying the case in which the number of bits used for representing each sample is fixed, and derive the sampling operator that minimizes the recovery mean-squared error (MSE). Then, we consider an overall bit budget, and optimize the number of bits assigned to each sample. We identify a sufficient condition for which the allocation is optimized by using identical quantizers, i.e., as in conventional task-based quantization [21].

The second choice of filter corresponds to frequency domain graph filters [8]. These filters are constrained to represent local computations, such that each node can be combined only with neighbouring nodes. Frequency domain graph filters notably facilitate distributed implementation, compared to unconstrained graph samplers which may require each sample to be produced by observing the entire graph [27]. For this constrained family, where the sampling operation is modeled by the selection of a subset of the outputs of a frequency domain graph filter, we first derive the sampling set and the corresponding assigned number of bits. Then we propose an alternating optimization algorithm to design the filter along with the overall compression system.

Our simulation study considers graph signal compression in three different applications: signals representing sensor networks; meteorological real-world data; and natural images. We consistently demonstrate that the proposed algorithm results in a distortion which is within a minor gap of that achievable without bit constraints, while yielding notably more accurate representations of the graph signal under a given bit budgetcompared to separately designed sampling and quantization.

The rest of the paper is organized as follows: Section II introduces the graph signal model, and formulates the compression problem. In Section III, the compression scheme is presented for generic graph filters, while frequency domain graph filters are considered in Section IV. Section V details numerical simulations. Finally, Section VI provides concluding remarks. Detailed proofs are delegated to the Appendix.

Throughout the paper, we use lower-case (upper-case) bold characters to denote vectors (matrices). The transpose, pseudo-inverse and trace of a matrix $\mathbf{A}$ are respectively denoted as $\mathbf{A}^T$, $\mathbf{A}^\dagger$ and Tr($\mathbf{A}$), while $(\mathbf{A})_{i,j}$ is the (i,j)th entry of $\mathbf{A}$, and $\mathbf{A}_\mathcal{S}$ represents a sub-matrix of $\mathbf{A}$ with rows indexed by $\mathcal{S}$. For a vector $\mathbf{a}$, $\mathbf{a}_i$ is its i-th element, and diag($\mathbf{a}$) is a diagonal matrix with $\mathbf{a}$ on its main diagonal. For a scalar a, $\lceil a \rceil$ and $\lfloor a \rfloor$ represent the round up and round down for a, respectively. We use $\mathbb{R}$ for the set of real numbers, $\mathbb{Z}$ for the integers, and $\mathbf{1}_\mathcal{A}$ is the indicator function for event $\mathcal{A}$.

II. System Model

In this section we present the system model for which we derive the joint sampling and quantization compression method. We first discuss the graph signal model in Subsection II-A, followed by a brief review of graph sampling and quantization in Subsection II-B, and a formulation of the considered problem in Subsection II-C.

A. Bandlimited Graph Signal Model

We consider an undirected and weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, in which $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes and $\mathcal{E}$ is the set of edges. Let $\mathbf{W} \in \mathbb{C}^{N\times N}$ be the adjacency matrix, such that its (i,j)th element $\mathbf{W}_{i,j}$ models the similarity/relationship between nodes i and j. A graph signal is a function $\mathbf{f} : \mathcal{V} \to \mathbb{R}^N$ defined on the vertices of $\mathcal{G}$. We adopt the symmetric normalized Laplacian matrix $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ as the variation operator, where $\mathbf{D} = \text{diag}\{d_1, d_2, \ldots, d_N\}$ is the degree matrix whose ith diagonal element is given by $d_i = \sum_j \mathbf{W}_{i,j}$. Since $\mathbf{L}$ is diagonalizable, there exists an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{\Lambda}$ satisfying $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. For a graph signal $\mathbf{x} \in \mathbb{R}^N$ defined over $\mathcal{G}$, its graph Fourier transform (GFT) is defined as $\hat{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$. The graph signal is bandlimited if there exists an integer $K \leq N$ such that its GFT satisfies $\hat{\mathbf{x}}_k = 0$ for all $k \geq K$. The smallest value of K denotes the bandwidth of $\mathbf{f}$. Bandlimited graph signals can be expressed as $\mathbf{U}_K\mathbf{c}$, where $\mathbf{U}_K$ is the first K columns of $\mathbf{U}$, while $\mathbf{c} \in \mathbb{R}^K$ is the frequency representation. Here, we consider a noisy observation of a bandlimited graph signal, given by

$$\mathbf{x} = \mathbf{U}_K\mathbf{c} + \mathbf{w}, \tag{1}$$

where $\mathbf{w}$ is an i.i.d. zero-mean noise with variance σ_0^2.

As in the classical factor analysis model [12], we impose a Gaussian prior on the spectral representation $\mathbf{c}$. Specifically, we assume that $\mathbf{c}$ follows a degenerate zero-mean multivariate Gaussian distribution such that $\mathbf{c} \sim \mathcal{N}(0, \hat{\mathbf{\Lambda}})$, where $\hat{\mathbf{\Lambda}} = \text{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\}$, and σ_i^2 represents the variance for i-th spectral representation $\mathbf{c}_i$ for $i \in \{1, \ldots, K\}$. Under this model, the signal $\mathbf{x}$ obeys a zero-mean multivariate Gaussian distribution with covariance matrix:

$$\mathbf{C}_\mathbf{x} = \mathbf{U}_K\hat{\mathbf{\Lambda}}\mathbf{U}_K^T + \sigma_0^2\mathbf{I} = \mathbf{U}\tilde{\mathbf{\Lambda}}\mathbf{U}^T, \tag{2}$$

where

$$\tilde{\mathbf{\Lambda}} = \text{diag}\left\{\sigma_1^2 + \sigma_0^2, \sigma_2^2 + \sigma_0^2, \ldots, \sigma_K^2 + \sigma_0^2, \sigma_0^2, \ldots, \sigma_0^2\right\}. \tag{3}$$

The joint Gaussianity of $\mathbf{c}$ and $\mathbf{x}$ implies that the minimum MSE (MMSE) estimate of the spectrum $\mathbf{c}$ from the noisy graph signal $\mathbf{x}$ is given by $\tilde{\mathbf{c}} = \mathbf{\Gamma}^*\mathbf{x}$, where

$$\mathbf{\Gamma}^* = \hat{\mathbf{\Lambda}}(\tilde{\mathbf{\Lambda}}_K)^{-1}\mathbf{U}_K^T. \tag{4}$$

Here, $\tilde{\mathbf{\Lambda}}_K$ represents a $K \times K$ matrix representing the first K rows and first K of $\tilde{\mathbf{\Lambda}}$. The resulting estimation error is given by

$$\mathbb{E}\{\|\tilde{\mathbf{c}} - \mathbf{c}\|^2\} = \text{Tr}\left(\hat{\mathbf{\Lambda}} - \mathbf{U}_K\tilde{\mathbf{\Lambda}}^2\mathbf{U}_K^T\mathbf{C}_{\mathbf{x}}^{-1}\right). \tag{5}$$

B. Sampling and Qauntization of Graph Signals

In this work, we combine sampling and quantization for compressing bandlimited graph signals. We thus next briefly recall some basics in graph signal sampling and quantization, starting with the definition of vertex-domain sampling:

Definition 1 (Vertex-domain sampling [8]): A vector $\mathbf{y} \in \mathbb{R}^P$ is the vertex-domain sampling of a graph signal $\mathbf{x} \in \mathbb{R}^N$, if there exists $\mathbf{\Psi} \in \mathbb{R}^{P \times N}$ such that $\mathbf{y} = \mathbf{\Psi}\mathbf{x}$ and $P < N$.

Vertex-domain sampling, abbreviated henceforth as sampling, often restricts $\mathbf{\Psi}$ to be a submatrix of the $N \times N$ identity matrix [6]. This results in the elements of $\mathbf{y}$ being also elements of $\mathbf{x}$. Nonetheless, graph sampling operations and their corresponding methods for reconstructing $\mathbf{x}$ from $\mathbf{y}$ are studied in the literature for various forms of $\mathbf{\Psi}$ [8]. Sampling matrices can be generally written as $\mathbf{\Psi} = \mathbf{I}_{\mathcal{S}}\mathbf{F}$, where $\mathbf{I}_{\mathcal{S}} \in \mathbb{R}^{P \times N}$ is a row-selection matrix, $\mathbf{F} \in \mathcal{F} \subseteq \mathbb{R}^{N \times N}$ is a graph filter, and $\mathcal{F}$ is the set of feasible graph filters.

We separately consider two forms of graph filters. The first type is referred to as unconstrained graph filters:

Definition 2 (Unconstrained graph filter): An unconstrained graph filter can be any $N \times N$ matrix, i.e., $\mathcal{F} = \mathbb{R}^{N \times N}$.

For unconstrained graph filters, the sampling matrix $\mathbf{\Psi}$ can be any matrix in $\mathbb{R}^{P \times N}$. The application of sampling using unconstrained graph filters requires the entire graph signal to be available for generating each sample, i.e., each element in the sampled $\mathbf{y}$ can be affected by every node in the graph. Such sampling procedures may be challenging to implement, particularly when dealing with high-dimensional graph signals. This motivates the restriction to graph sampling procedures which involve only local operations, where each sample is affected only by a subset of neighbouring nodes. Such sampling mechanisms can be realized by constraining $\mathbf{F}$ to represent *frequency-domain filtering*, defined next:

Definition 3 (Frequency-domain graph filter [7]): The set of frequency-domain graph filters defined over a graph $\mathcal{G}$ with Laplacian $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ is given by

$$\mathcal{F} = \{\mathbf{F} = \mathbf{U}F(\mathbf{\Lambda})\mathbf{U}^T | F(\mathbf{\Lambda}) \text{ is diagonal}\}. \tag{6}$$

Similarly to unconstrained graph filters in Definition 2, frequency-domain graph filters are also linear operators, and are thus represented via a sampling matrix $\mathbf{\Psi} \in \mathbb{R}^{P \times N}$. The constraint to represent a frequency-domain graph filter indicates that $\mathbf{\Psi}$ should be of the form $\mathbf{\Psi} = \mathbf{I}_{\mathcal{S}}\mathbf{U}F(\mathbf{\Lambda})$, and thus the sampled vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{I}_{\mathcal{S}}\mathbf{U}F(\mathbf{\Lambda})\hat{\mathbf{x}}. \tag{7}$$

The representation of the graph sampling operation via (7) notably facilitates distributed operation compared with sampling using unconstrained graph filters. This follows since (7) can be approximated by interpolation [28], resulting in the frequency-domain graph filter being expressed as a matrix polynomial function with respect to $\mathbf{L}$. In particular, a frequency domain graph filter restricted to combining elements corresponding the neighbouring nodes of the graph implies that $F(\mathbf{\Lambda})$ should be a polynomial function with respect to $\mathbf{\Lambda}$, i.e., $F(\mathbf{\Lambda}) = \sum_{i=0}^{K_0}\beta_i\mathbf{\Lambda}^i$, for some coefficients $\{\beta_i\}$, where K_0 is the length of the polynomial [27].

Next, we discuss the considered quantization rule. While in general, quantization encapsulates a broad range of continuous-to-discrete mappings [15], here we focus on conventional uniform scalar quantizaiton, defined as follows:

Definition 4 (Uniform quantization): A uniform quantizer with resolution M, support γ, and step size $\delta = \frac{2\gamma}{M}$, is the mapping

$$Q_M(x) \triangleq \begin{cases} \delta\left(\lfloor\frac{x}{\delta}\rfloor + \frac{1}{2}\right), & \text{if } |x| < \gamma, \\ \text{sign}(x)\left(\gamma - \frac{\delta}{2}\right), & \text{else.} \end{cases}$$

The non-linear nature of quantization mappings results in a complex model for the distortion induced in this procedure, i.e., the term $x - Q_M(x)$. In our analysis we model the quantization operation as *non-subtractive dithered quantization*, obtained by adding noise of a triangular distribution over $[-\delta, \delta]$ prior to uniform quantization [24], [29]. The main motivation for this model is that it results in the distortion being uncorrelated with the input signal when the quantizer is not overloaded, i.e., $|x| \leq \gamma$. This model, which notably facilitates the design and analysis of quantization systems, is also a good approximation of the distortion model for various quantizer input distributions without dithering as analyzed in [30] as demonstrated in [21].

C. Problem Formulation

We consider the problem of compressing a noisy graph signal $\mathbf{x}$ into a digital representation comprised of $\log_2 M$ bits. This compressed version should preserve the information required to reconstruct the bandlimited graph signal $\mathbf{U}_K\mathbf{c}$. We assume that the structure of the graph signal and its spectral support are known, i.e., $\mathbf{U}_K$ is given.

Since the volume of the representation is limited in its overall number of bits, we study compression mechanisms consisting of both sampling and quantization, as defined in Subsection II-B. In such a system, the graph signal is first sampled by applying the $P \times N$ matrix $\mathbf{\Psi}$, with $P < N$, and the entries of the sampled $\mathbf{y} = \mathbf{\Psi}\mathbf{x}$ are then discretized by the uniform quantizers $\{Q_{M_i}(\cdot)\}_{i=1}^P$, resulting in the finite-bit representation $Q(\mathbf{y}) = [Q_{M_1}(y_1), \ldots, Q_{M_P}(y_P)]^T$. The overall bit constraint implies that $\sum_{i=1}^P \log_2 M_i \leq \log_2 M$.

As quantization inherently results in lossy compression [31], Ch. 23], we do not aim to perfectly recover $\mathbf{x}$, and use the MSE as our design objective. Consequently, the joint design of the sampling and quantization mappings can be formulated as recovering the digital representation from which the bandlimited portion of $\mathbf{x}$, i.e., $\mathbf{U}_K\mathbf{c}$, can be reconstructed most accurately.

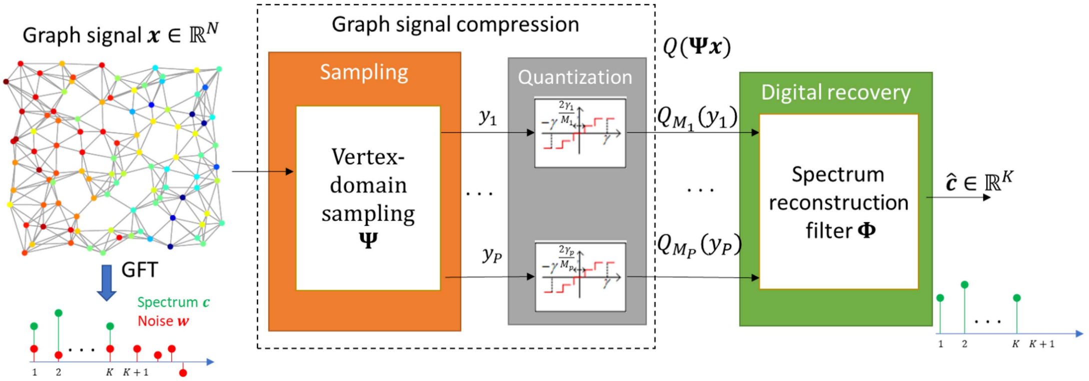

Graph signal $\boldsymbol{x} \in \mathbb{R}^N$
GFT
Spectrum $\boldsymbol{c}$
Noise $\boldsymbol{w}$
1 2 . . . K K+1
Graph signal compression
Sampling
Vertex-domain sampling $\boldsymbol{\Psi}$
y_1
. . .
y_P
Quantization
$Q(\boldsymbol{\Psi x})$
$Q_{M_1}(y_1)$
. . .
$Q_{M_P}(y_P)$
Digital recovery
Spectrum reconstruction filter $\boldsymbol{\Phi}$
$\hat{\boldsymbol{c}} \in \mathbb{R}^K$
1 2 . . . K K+1

Fig. 1. Graph signal compression by sampling and quantization.

This is mathematically formulated as

$$\min_{\boldsymbol{\Psi}, Q(\bullet)} \mathbb{E}\left\{\left\|\mathbf{U}_K\mathbf{c} - \mathbb{E}\left\{\mathbf{U}_K\mathbf{c}|Q(\boldsymbol{\Psi}\mathbf{x})\right\}\right\|^2\right\},$$
$$\text{s.t.} \sum_{i=1}^{P} \log_2 M_i \leq \log_2 M, \quad M_i \in \mathbb{Z}^+, i \in \mathcal{P}, \qquad \text{(P1)}$$

where $\mathcal{P} \triangleq \{1, 2, \ldots, P\}$. In the following sections we jointly design the resulting compression system based on (P1).

III. Joint Sampling and Qauntization Method With Unconstrained Graph Filters

In this section we derive joint sampling and quantization methods for graph signal compression with unconstrained graph filters. To that aim, we first present in Subsection III-A an alternative problem formulation, obtained by introducing some relaxations to (P1). Then, we present the MSE minimizing system design in Subsection III-B, and provide a greedy-based algorithm to tackle it in Subsection III-C.

A. Alternative Optimization Problem

The problem formulation (P1) characterizes the graph signal compression setup as designing the sampling matrix $\boldsymbol{\Psi}$ and the scalar quantizers $Q(\bullet)$. The resulting setup bears much similarity to task-based quantization, proposed as a framework for designing ADCs to extract information from an acquired analog signal [21], [22], [23], [24]. To exploit task-based quantization in graph signal compression, we formulate an alternative problem based on (P1), which can be tackled using these techniques.

First, we note that since the unitary matrix $\mathbf{U}$ and its submatrix $\mathbf{U}_K$ are known, recovering the MMSE estimate of $\mathbf{x}$, as formulated in (P1), is equivalent to the corresponding estimation of $\mathbf{c}$. We thus henceforth refer to $\mathbf{c}$ as the *task vector*. Furthermore, recalling the definition of $\tilde{\mathbf{c}} = \mathbb{E}\{\mathbf{c}|\mathbf{x}\} = \boldsymbol{\Gamma}^*\mathbf{x}$, then by the orthogonality principle, designing $Q(\bullet)$ and $\boldsymbol{\Psi}$ to minimize the MSE w.r.t. $\mathbf{c}$ is the same as designing them to minimize the MSE w.r.t. $\tilde{\mathbf{c}}$. The objective in (P1) thus becomes

$$\min_{\boldsymbol{\Psi}, Q(\bullet)} \mathbb{E}\left\{\left\|\tilde{\mathbf{c}} - \mathbb{E}\left\{\tilde{\mathbf{c}}|Q(\boldsymbol{\Psi}\mathbf{x})\right\}\right\|^2\right\}. \qquad (8)$$

Next, we relax (8) by focusing on linear recovery. Our motivation for considering linear schemes stems from their analytical tractability, and since they are commonly used for reconstructing sampled graph signals [8]. The compression system is designed such that the desired vector $\mathbf{c}$ can be recovered from $Q(\boldsymbol{\Psi}\mathbf{x})$ using a filter $\boldsymbol{\Phi} \in \mathbb{R}^{K \times P}$, i.e., the recovered $\mathbf{c}$ is given by $\hat{\mathbf{c}} = \boldsymbol{\Phi} Q(\mathbf{y})$, where $\mathbf{y} = \boldsymbol{\Psi}\mathbf{x}$. This compression and recovery system is illustrated in Fig. 1.

Quantizers are typically designed to operate within their dynamic range [15]. Therefore, following [21], we guarantee that the probability of overloading the quantizers is sufficiently small, i.e., that $\Pr(|(\boldsymbol{\Psi}\mathbf{x})_i| > \gamma_i) \approx 0$, by fixing the support of the ith quantizer to $\gamma_i^2 = \eta^2\mathbb{E}\{(\boldsymbol{\Psi}\mathbf{x})_i^2\}$ for each $i \in \mathcal{P}$. For instance, setting $\eta = 2$ results in overloading probability $< 5\%$. When the overloading probability vanishes, then the output of the dithered quantizers can be written as [29]

$$Q\left(\boldsymbol{\Psi}\mathbf{x}\right) = \boldsymbol{\Psi}\mathbf{x} + \mathbf{e}_Q, \qquad (9)$$

where the quantization error vector $\mathbf{e}_Q$ has i.i.d. zero-mean entries of variance $\delta_i^2/4$ (with $\delta_i = 2\gamma_i/M_i$), i.e.,

$$\mathbf{G} \triangleq \mathbb{E}\left\{\mathbf{e}_Q\mathbf{e}_Q^T\right\} = \operatorname{diag}\left\{\frac{\gamma_1^2}{M_1^2}, \frac{\gamma_2^2}{M_2^2}, \ldots, \frac{\gamma_P^2}{M_P^2}\right\}. \qquad (10)$$

While our setting of γ_i^2 achieves a small yet non-zero overloading probability, we design the compression mechanism by utilizing the model in (9), which rigorously holds when the overloading probability is zero, as proposed in [21], [22].

To summarize, the alternative optimization problem based on which we design our scheme is given by

$$\min_{\boldsymbol{\Psi}, \boldsymbol{\Phi}, \{M_i\}} \mathbb{E}\left\{\left\|\boldsymbol{\Gamma}^*\mathbf{x} - \boldsymbol{\Phi}(\boldsymbol{\Psi}\mathbf{x} + \mathbf{e}_Q)\right\|^2\right\},$$
$$\text{s.t.} \sum_{i=1}^{P} \log_2 M_i \leq \log_2 M, \quad M_i \in \mathbb{Z}^+, i \in \mathcal{P}. \qquad \text{(P2)}$$

Problem (P2) constitutes a simplification and a relaxation of the original (P1), and is equivalent to (P1) when the dithered quantizers are not overloaded. A key benefit of using the relaxed optimization problem (P2) stems from the fact that it can be treated as a task-based quantization problem. In task-based quantization, an analog filter is designed along with the overall acquisition system in light of the system task [21]. Here, the sampling operation implements the pre-quantization combining, and the system task is to recover the vector $\mathbf{c}$ (via its MMSE estimate). However, as task-based quantization focuses on the design of ADCs operating in a serial manner, it is restricted to utilizing identical quantizers, i.e., each quantizer uses the same number of bits, while our graph signal compression problem does not share this restriction. Therefore, our design based on (P2), given in the following subsection, combines task-based quantization methods with dedicated bit allocation techniques.

B. Compression System Design

In this section, we design a joint graph signal sampling and quantization scheme based on the identified relationship between such setups and task-based quantization in (P2). Directly solving (P2) is difficult due to the coupling between its optimization variables, combined with the non-linear relationship between these parameters and the statistical model of the quantization error, observed in (10). Therefore, in the following, we first design the sampling matrix based on (P2) for a fixed bit allocation, i.e., when $\{M_i\}$ is given. To motivate our quantization-aware design, we first discuss the intuitive setting of frequency-domain sampling.

Example 1: A natural approach to sample bandlimited graph signals is to set the sampling matrix to be $\mathbf{\Psi} = \mathbf{U}_K^T$ (where if $P > K$, it is padded with zeros, effectively using K quantizers). This approach samples by filtering the frequency components of the noise $\mathbf{w}$ in (1) that are outside the signal frequency support. Using (9), the compressed signal here can be expressed as

$$Q\left(\mathbf{\Psi x}\right) = \mathbf{c} + \mathbf{U}_K^T \mathbf{w} + \mathbf{e}_Q. \tag{11}$$

The compressed signal in (11) is given by the desired spectral components $\mathbf{c}$ corrupted by a distortion term $\mathbf{U}_K^T\mathbf{w} + \mathbf{e}_Q$ that is uncorrelated with $\mathbf{c}$. Further, this distortion term is comprised of uncorrelated entries, where the variance of its ith entry by (10) is

$$\sigma_0^2 + \frac{\gamma_i^2}{M_i^2} = \sigma_0^2 + \frac{\eta^2(\sigma_i^2 + \sigma_0^2)}{M_i^2}. \tag{12}$$

The representation of the compressed signal for the intuitive setting of frequency domain sampling gives rise to three main insights, that motivate our design in the sequel:

1) Even with infinite resolution quantization, i.e., when $M_i \to \infty$, simple frequency domain sampling still requires some linear processing to reduce the recovery MSE, i.e., (11) needs to be multiplied by $\hat{\mathbf{\Lambda}}(\tilde{\mathbf{\Lambda}}^{-1})_K$ to minimize the MSE by (4);
2) The equivalent distortion in each entry of $Q(\mathbf{\Psi x})$ dependeds not only of the bit allocation $\{M_i\}$, but also on the energy of the corresponding sample. This indicates that for a given bit allocation, one can possibly identify sampling matrices which yield a distortion profile that better facilitates recovery compared to that in (12) achieved when $\mathbf{\Psi} = \mathbf{U}_K^T$;
3) The ability to tune the bit allocation $\{M_i\}$ indicates that one can control how the distortion is distributed among the samples, e.g., (12). However, since $\{M_i\}$ must be positive integers, not every distortion profile is achieved, and thus the sampling matrix $\mathbf{\Psi}$ should also be tuned and possibly deviate from the intuitive setting of $\mathbf{\Psi} = \mathbf{U}_K^T$.

Example 1 indicates that additional processing of the compressed signal is required to improve reconstruction of the spectrum. We note that for any given sampling matrix $\mathbf{\Psi}$ and bit allocation $\{M_i\}$, the MSE minimizing linear recovery matrix $\mathbf{\Phi}$ is given in the following lemma:

Lemma 1: For any fixed $\mathbf{\Psi}$ and $\{M_i\}$, the objective in (P2) is minimized by setting

$$\mathbf{\Phi}^* = \mathbf{\Gamma}^* \mathbf{C}_{\mathbf{x}} \mathbf{\Psi}^T \left(\mathbf{\Psi} \mathbf{C}_{\mathbf{x}} \mathbf{\Psi}^T + \mathbf{G}\right)^{-1}. \tag{13}$$

Proof: The lemma is obtained following [21, App. B]. ■

Next, we design the sampling matrix $\mathbf{\Psi}$. Here, we recall the similarity between our sampling matrix and analog combiner design in task-based quantization setups. In particular, the MSE minimizing analog combiner for task-based quantization is given in [21, Thm. 1] for an identical bit allocation $\tilde{M} = \lfloor M^{1/P} \rfloor$. For such setups, the MSE is minimized for a setting of the form $\mathbf{\Psi} = \mathbf{U \Xi V}^T$, where $\mathbf{V} \in \mathbb{R}^{N \times N}$ is the right singular vectors matrix of $\mathbf{\Gamma}^* \mathbf{C}_{\mathbf{x}}^{1/2}$, $\mathbf{\Xi} \in \mathbb{R}^{P \times N}$ is diagonal with non-negative diagonal entries, and $\mathbf{U} \in \mathbb{R}^{P \times P}$ is a unitary matrix designed to balance the inputs to the identical quantizers. Since in our setup, the quantizers are not restricted to be identical, we cannot adopt the results derived in [21]. Nonetheless, for a given bit allocation $\{M_i\}$ sorted in a descending order, we can characterize the MSE minimizing sampling matrix, as stated in the following proposition:

Proposition 1: For a bit allocation $\{M_i\}$ with $M_i \geq M_{i+1}$, the MSE minimizing unconstrained sampling matrix is given by

$$\mathbf{\Psi}^* = \mathbf{U}_\Psi \mathbf{\Xi}_\Psi \tilde{\mathbf{\Lambda}}^{-1/2} \mathbf{U}^T, \tag{14}$$

where $\mathbf{U}_\Psi$ is a unitary matrix satisfying

$$(\mathbf{U}_\Psi \mathbf{\Xi}_\Psi \mathbf{\Xi}_\Psi^T \mathbf{U}_\Psi^T)_{i,i} = \frac{M_i^2}{\eta^2}, \tag{15}$$

and $\mathbf{\Xi}_\Psi \in \mathbb{R}^{P \times N}$ is a diagonal matrix with non-negative entries. Denoting $\alpha_i = (\mathbf{\Xi}_\Psi)_{i,i}^2, i \in \mathcal{P}$, the elements of $\mathbf{\Xi}_\Psi$ are obtained as the solution to the following problem:

$$\{\alpha_i\}_{i=1}^P = \underset{\{\alpha_i'\}}{\arg\min} \sum_{i=1}^P \frac{\lambda_{\mathbf{\Gamma},i}^2}{\alpha_i' + 1},$$
$$\text{s.t. } \frac{1}{\eta^2}[M_1^2, \ldots M_P^2]^T \prec [\alpha_1', \ldots \alpha_P']^T, \tag{16}$$

where $\prec$ denotes majorization, i.e., $\boldsymbol{a} \prec \boldsymbol{b}$ implies that $\boldsymbol{a}$ is majorized by $\boldsymbol{b}$ [32]. In (16), $\lambda_{\mathbf{\Gamma},i}$ is the ith largest singular value of $\mathbf{\Gamma}^* \mathbf{C}_{\mathbf{x}}^{1/2}$. The resulting excess MSE is given by:

$$\mathbb{E}\{\|\hat{\mathbf{c}} - \tilde{\mathbf{c}}\|^2\} = \sum_{i=1}^P \frac{\lambda_{\mathbf{\Gamma},i}^2}{\alpha_i + 1}. \tag{17}$$

Proof: The proof is given in Appendix A. ■

For the special case of identical bit allocation, i.e., $M_i = M_j$ for each $i, j \in \mathcal{P}$, it can be shown that Proposition 1 coincides with [21, Thm. 1]. We note that (16) is a convex problem, which can be solved using existing convex optimization toolboxes such as CVX. Consequently, while Proposition 1 does not give the sampling matrix $\mathbf{\Psi}^*$ in closed form, it can be numerically computed with an affordable computational effort. Nonetheless, identifying the bit allocation $\{M_i\}$ which minimizes the MSE based on Proposition 1 is a challenging task due to the discrete nature of the bit assignment, motivating the greedy optimization algorithm detailed in the following section. However, if one is allowed to assign non-integer bit values, the resulting compression system and its corresponding MSE can be obtained explicitly via the following theorem:

Theorem 1: When $\{M_i\}$ are not limited to be integer, the optimal unconstrained sampling operator is $\mathbf{\Psi}^* = \mathbf{U}_K^T$. The MSE minimizing bit allocations satisfies

$$M_i^2 = \begin{cases} \eta^{2\frac{-\beta^* + \lambda^2_{\mathbf{\Gamma},i} + \sqrt{\lambda^4_{\mathbf{\Gamma},i} - \beta^* \lambda^2_{\mathbf{\Gamma},i}}}{\beta^*}} & \beta^* \leq \frac{\lambda^2_{\mathbf{\Gamma},i}}{4}, \\ 1 & \text{otherwise}, \end{cases} \tag{18}$$

where the hyperparameter β^* is set such that $\sum_{i=1}^P \log_2 M_i = \log_2 M$. The resulting excess MSE is given by:

$$\mathbb{E}\{\|\hat{\mathbf{c}} - \tilde{\mathbf{c}}\|^2\} = \sum_{i=1}^{K} \lambda^2_{\tilde{\Gamma},i} - \sum_{i=1}^{\min(K,P)} \frac{M_i^2 \lambda^2_{\tilde{\Gamma},i}}{M_i^2 + \eta^2}. \tag{19}$$

Proof: The proof is given in Appendix B. ■

While Theorem 1 considers a hypothetical system which can quantize using non-integer number of bits, it reveals an important challenge arising from the incorporation of quantization compared to conventional graph sampling. When one can quantize with arbitrary non-integer levels, the resulting compression problem reduces to a conventional graph sampling (without quantization) setup, as shown in Appendix B. For such cases, the sampling matrix $\mathbf{\Psi}^*$ specializes to conventional frequency domain sampling of bandlimited signals. This settles with corresponding results in [8], as well as with the observation in Example 1, where we note that when using frequency domain sampling, one can further tune how the distortion is distributed among the samples via the bit allocation $\{M_i\}$. However, since quantization is inherently restricted to discrete levels, the MSE minimizing graph sampling matrix does not necessarily reduce to frequency domain sampling, as the equivalent distortion cannot be controlled to any desired resolution. In such cases, one has to utilize Proposition 1 combined with dedicated schemes for just rounded or optimizing the bit allocation, as proposed next.

C. Greedy Optimization Algorithm Design

The MSE characterized Theorem 1 is generally not achievable for graph signal compression schemes since it ignores the fact that the bit assignment must take integer values. Nonetheless, the closed-form MSE expression (19) can be used to facilitate the setting of the (discrete) bit allocation $\{M_i\}$ using greedy optimization. In the following we first introduce the greedy algorithm for setting the bit budget, after which we identify a sufficient and necessary conditions for which the bit assignment is designed by using identical quantizers, and identify the number of quantizers P which minimizes the MSE without setting any quantizer to be inactive.

Algorithm 1: Greedy Bit Allocation Algorithm.

```
Input: $\{\lambda_{\tilde{\Gamma},i}\}$
Output: Bit allocation $\{M_i\}$.
Initialize: $k = 1$ and set $M_i^{(0)} = 1$ for $\forall i \in \mathcal{P}$.
while $\sum_{i=1}^P \log_2 M_i^{(k)} \leq \log_2 M$ do
    Compute $\mathbf{g}^{(k)}$ via (20);
    Update $M_i^{(k+1)} = M_i^{(k)} + \mathbf{1}_{i=\arg\min_j g_j(M_j^{(k)})}$;
    $k = k + 1$;
end while
return $\{M_i^{(k-1)}\}$
```

1) Greedy Bit Assignment: The proposed greedy algorithm starts by setting the minimal bit allocation for all the quantizers, i.e., $M_i^{(0)} = 1$ for each $i \in \mathcal{P}$. Then, it increments the number of levels for the quantizer which contributes most to the objective (19). While this procedure does not impose the constraint $M_i \geq M_{i+1}$ for each $i \in \{1, 2, \ldots, P-1\}$, it is implicitly maintained due to the descending order of $\{\lambda_{\mathbf{\Gamma},i}\}$. Specifically, the gradient of the objective (19) w.r.t. M_i is $g(M_i) := -\frac{\partial}{\partial M_i} \sum_{i=1}^{\min(K,P)} \frac{M_i^2 \lambda^2_{\tilde{\Gamma},i}}{M_i^2 + \eta^2}$, which equals

$$g_i(M_i) = \begin{cases} -\frac{M_i \eta^2 \lambda^2_{\mathbf{\Gamma},i}}{(M_i^2 + \eta^2)^2} & i \leq K, \\ 0 & \text{otherwise}. \end{cases}$$

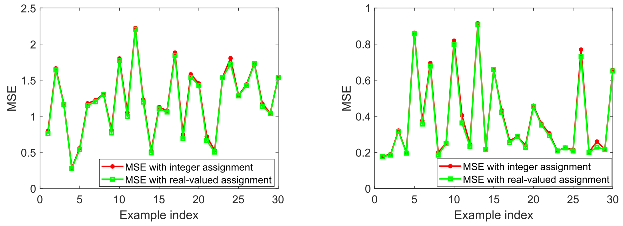
At the kth iteration, the gradient vector $\mathbf{g}^{(k)}$ is thus given by

$$\mathbf{g}^{(k)} = [g_1(M_1^{(k)}), g_2(M_2^{(k)}), \ldots, g_P(M_P^{(k)})]^T. \tag{20}$$

The elements in (20) respectively represent the distortion decreasing efficiency of the extra level we assign to each quantizer at the kth iteration. The smallest entry of (20) determines which quantizer is assigned an additional level, repeating until the budget M is exhausted. The resulting greedy allocation algorithm is summarized as Algorithm 1. Once the bit assignment is obtained, the sampling operator and its corresponding MSE are obtained using Proposition 1, while the digital reconstruction filter is computed via Lemma 1.

2) Analysis of Algorithm 1: In the method summarized as Algorithm 1, we use a greedy approach based on an MSE measure achievable using real-valued assignments to optimize the discrete bit allocation. Therefore, to assess the validity of this approach, we first numerically compare the ability of Algorithm 1 to approach the excess MSE in (19) using discrete bit assignments. Then, we characterize a sufficient condition for which Algorithm 1 yields an identical bit allocation. Since the number of quantized samples P is inherently a design parameter of the system, we derive the setting of P which optimizes the MSE achievable using Algorithm 1. Finally, we analyze the computational complexity of Algorithm 1.

Empirical Evaluation: Fig. 2 compares the MSE achieved with the discrete bit assignment computed using Algorithm 1 to the MSE evaluated via (19) for 30 i.i.d. examples with $P = K = 10$. As the purpose of this study is to evaluate the ability of Algorithm 1 to approach (19), we consider a toy example



(a) $\log_2 M = 10$ (b) $\log_2 M = 20$

Fig. 2. 30 examples with randomly generated $\lambda_{\Gamma,i}$.

where the descending sequence $\{\lambda_{\tilde{\Gamma},i}\}$ is randomly generated with normal Gaussian distribution; more realistic experiments corresponding to different forms of graph signals are reported in our simulation study in Section V.

Observing Fig. 2, we note that when the overall bit budget is tight, i.e., $\log_2 M = 10$ bits as in Fig. 2(a), there is a small gap between the MSE achieved using the discrete assignment and that which can reached given the ability to assign arbitrary quantization levels. However, as the bit budget increases to $\log_2 M = 20$, which is still a relatively tight budget (allowing merely 2 bits per quantizer for a standard identical bit assignment), we observe in Fig. 2(b) that the gap becomes negligible. This study indicates that although Algorithm 1 is designed based on an MSE measure typically not achievable with integer bit assignments, it allows to approach it to within a small gap, which effectively vanishes as the bit budget grows.

Identical Bit Assignment: As discussed in Subsection III-B, while our derivation relies on representing the joint sampling and quantization of bandlimited graph signals as a task-based quantization setup, the resulting formulation differs from that studied in the context of ADC design in [21]. One of the key differences between the graph signal compression task considered here and the design of hybrid analog/digital acquisition studied in [21] stems from the additional degree of freedom in the ability to allocate different bit assignments between the quantizers. As the restriction to utilize identical quantizers may facilitate the design of the compression system, we next identify a sufficient and necessary condition for which Algorithm 1 yields an identical bit allocation:

Corollary 1: Algorithm 1 yields an identical bit assignment $M_i \equiv M_a := \lfloor M^{1/P} \rfloor$ for each $i = 1, \ldots, P$ where $P \leq K$ when the following conditions are satisfied:

$$(M_a)^P \leq M < (M_a)^P + (M_a)^{P-1}, \tag{21a}$$

$$g_1(M_a) < g_P(M_a - 1). \tag{21b}$$

Proof: The proof is given in Appendix C. ■

Corollary 1 identifies conditions for which, when satisfied, one can simply utilize conventional identical bit allocation rather than go through the greedy optimization in Algorithm 1. Nonetheless, the conditions (21) are not that commonly satisfied by graph signals. The condition (21b) hints that the $\lambda_{\Gamma,P}$ should be close to $\lambda_{\Gamma,1}$. However, the smoothness of the graph signal determines that the weight of the low-frequency component is usually much greater than the weight of the high-frequency component, i.e., $\sigma_1 \gg \sigma_P$, which potentially contradicts (21b). On the other hand, when (21a) is not satisfied, the greedy-based algorithm we proposed can also improve the performance by using the redundant bits.

Number of Quantized Samples: Algorithm 1 assigns the overall bit budget among P quantized samples, where P is fixed. Nonetheless, it may yield an assignment $M_i = 1$, implying that the ith quantizer has a single quantization level and is thus inactive. For instance, since the gradients in (20) equal zero for $i > K$ and are strictly negative for $i \leq K$, it follows that Algorithm 1 assigns zero bits (i.e., $M_i = 1$) for quantizers of index $i > K$. While increasing P to be larger than the spectral support K thus does not affect the MSE as these quantizers will no be active, using less than K quantized samples may result in different achievable MSE values. Consequently, in the following corollary we identify the number of quantized samples P which minimizes the MSE without setting any quantizer to be inactive:

Corollary 2: The MSE minimizing number of actively quantized samples P^* satisfies $l_{P^*} \leq \log_2 M < l_{P^*+1}$, where

$$l_i = \begin{cases} 1 + \sum_{j=1}^{i} \log_2 \lceil \tilde{l}_{j,i} \rceil, & i \leq K, \\ +\infty, & i > K. \end{cases} \tag{22}$$

Here, $\tilde{l}_{j,i} \in \mathbb{R}$ is given by the solution to the following equality:

$$g_j(\tilde{l}_{j,i}) = g_i(1), \tag{23}$$

where $g_i(\bullet)$ is i-th element in the gradient of (19) w.r.t. M_i.

Proof: The proof is given in Appendix D. ■

Corollary 2 provides an increasing sequence, which divides the feasible interval of M decision levels into K sub-intervals. Before designing the graph signal compression system, we can determine the expected number of quantizers according to the size of M. When the total number of bits is large enough satisfying $\log_2 M \geq l_K$, K samples are actively quantized; When the total number of bits is limited, as discussed above, some inactive quantizers can be removed. In such a case, the reduction in the number of quantized samples also reduces the computational complexity of Algorithm 1 since it needs to compute P-length gradient vector in each iteration. However, for a specific graph signal compression task, if the number of samples that can be taken P is less than P^* of Corollary 2, then all samples should be quantized with at least two bits.

Complexity Analysis: To characterize the complexity order of Algorithm 1, we first focus on the complexity in each iteration. Each iteration is comprised of evaluating $\mathbf{g}^{(k)}$ (Step 2) and the update of M_i (Step 3). Since the computation of each element of $\mathbf{g}^{(k)}$ via (20) is fixed and does not grow with the system parameters, the complexity order of Step 2 is $\mathcal{O}(P)$. Updating M_i involves identifying the smallest entry of the $P \times 1$ vector $\mathbf{g}^{(k)}$, and thus its complexity order is at most $\mathcal{O}(P)$. Thus the complexity in each iteration is $\mathcal{O}(P)$, and since the The maximum number of iterations of Algorithm 1 is M, the overall complexity order of Algorithm 1 is $\mathcal{O}(MP)$.

IV. Optimization of Sampling and Quantization Using Frequency-Domain Graph Filters

Section III designs the compression rule without imposing any constraints on the sampling matrix, i.e., setting $\mathbf{\Psi} = \mathbf{I}_{\mathcal{S}}\mathbf{F}$, thus allowing the graph filter $\mathbf{F}$ to be any $N \times N$ matrix. In this section, we specialize our analysis to frequency-domain graph filters. Since we consider bandlimited graph signals, we henceforth focus on frequency domain graph filters applied in the spectral support of the signal, i.e., filters of the $\mathbf{F} = \mathbf{U}_K F(\mathbf{\Lambda})\mathbf{U}_K^T$, with $F(\mathbf{\Lambda})$ being a $K \times K$ diagonal matrix. We first present in Subsection IV-A an alternative problem formulation, obtained by restricting the graph filter in (P2) to implement frequency-domain graph filtering. Based on the modified problem formulation, we propose an algorithm to tune the bit allocation and sampling set design in Subsection IV-B, after which we provide an alternating optimization algorithm to obtain the overall compression scheme in Subsection IV-C, and discuss the resulting design in Subsection IV-D.

A. Problem Formulation

To formulate the compression problem using frequency-domain graph filters, we return to (P2), which serves as a basis for the design of the unconstrained system in Section III. The resulting formulation is stated in the following lemma:

Lemma 2: When the sampling matrix $\mathbf{\Psi}$ is restricted to represent frequency-domain graph filtering, i.e., $\mathbf{\Psi} = \mathbf{I}_{\mathcal{S}}\mathbf{U}_K F(\mathbf{\Lambda})\mathbf{U}_K^T$, then, by defining $\mathbf{H} \triangleq \mathbf{U}_K F(\mathbf{\Lambda})\tilde{\mathbf{\Lambda}}_K^2 F(\mathbf{\Lambda})\mathbf{U}_K^T$ and $\mathbf{X} \triangleq \mathbf{U}_K F(\mathbf{\Lambda})\tilde{\mathbf{\Lambda}}_K F(\mathbf{\Lambda})\mathbf{U}_K^T$, (P2) is specialized into

$$\max_{\mathbf{I}_{\mathcal{S}}, F(\mathbf{\Lambda}), \{M_i\}} \operatorname{Tr}\left(\mathbf{I}_{\mathcal{S}}\mathbf{H}\mathbf{I}_{\mathcal{S}}^T\left(\mathbf{I}_{\mathcal{S}}\mathbf{X}\mathbf{I}_{\mathcal{S}}^T + \mathbf{G}\right)^{-1}\right),$$
$$\text{s.t.} \sum_{i=1}^{P} \log_2 M_i \leq \log_2 M,\ M_i \in \mathbb{Z}^+, i \in \mathcal{P}. \quad \text{(P3)}$$

Proof: The proof is given in Appendix E. ■

Problem (P3) specializes (P2) to sampling using frequency-domain graph filtering. It translates the compression system design into the joint optimization of the frequency-domain graph filter $\mathbf{F} = \mathbf{U}_K F(\mathbf{\Lambda})\mathbf{U}_K^T$, the sampling set selection $\mathbf{I}_{\mathcal{S}}$, and the bit allocation $\{M_i\}$. In particular, $F(\mathbf{\Lambda})$ affects the matrices $\mathbf{H}$ and $\mathbf{X}$, while the bit setting is implicitly encapsulated in the distortion matrix $\mathbf{G}$ (10). Problem (P3) is non-convex due to the coupling of its variables, i.e., the fact that its objective includes products of its multiple different optimization variables. To tackle this challenging optimization, in the following subsections, we first show how one can tune $\mathbf{I}_{\mathcal{S}}$ and $\{M_i\}$ for a given graph filter $\mathbf{F}$, based on which we propose an alternating optimization algorithm to set the compression system parameters.

B. Sampling Set and Bit Allocation Design

Here, we solve (P3) under a given filter $\mathbf{F}$ to obtain the sampling set $\mathcal{S}$ and bit allocation $\{M_i\}$. While the number of samples taken is $|\mathcal{S}| = P$, the number of actively quantized samples can be smaller, as revealed by Corollary 2. Consequently, while the number of samples P, is fixed here, the optimization of the sampling set should account for the fact that possibly less than P samples are actively quantized. This requirement was not encountered when considering unconstrained graph filters, where $\mathbf{\Psi}$ can be any $P \times N$ matrix, while here $\mathbf{\Psi}$ needs to be explicitly divided into sampling set selection $\mathbf{I}_{\mathcal{S}}$ and the graph filter $\mathbf{F}$.

To optimize the sampling set along with the bit allocation $\{M_i\}$ in light of the above consideration, we allow $\mathcal{S}$ to contain only the actively quantized, i.e., $|\mathcal{S}| \leq P$. We can now define the following optimization variables

$$\tilde{M}_i \triangleq \begin{cases} M_j & i = (\mathcal{S})_j, \\ 1 & \text{otherwise}, \end{cases} \tag{24}$$

where $(\mathcal{S})_j$ is the j-th sampling vertex in $\mathcal{S}$. Note that in (24) there is a one-to-one correspondence between $\{M_j\}$, $\mathcal{S}$ and $\{\tilde{M}_i\}$. In particular, when $\tilde{M}_i = 1$, then the ith node of the graph signal at the output of the graph filter $\boldsymbol{F}$ is quantized with zero bits, i.e., its value is ignored, and it is thus not sampled. Consequently, $\{M_j\}$ and $\mathcal{S}$ are uniquely determined by $\{\tilde{M}_i\}$. Therefore, by letting $\mathbf{G}_{\{\tilde{M}_j\}}$ be the matrix $\mathbf{G}$ in (10) with the bit allocation n $\{\tilde{M}_j\}$, the optimization of the sampling set and the bit allocation for a given $\mathbf{F}$ can be formulated as

$$\max_{\{\tilde{M}_i\}} \operatorname{Tr}\left(\mathbf{I}_{\mathcal{S}}\mathbf{H}\mathbf{I}_{\mathcal{S}}^T\left(\mathbf{I}_{\mathcal{S}}\mathbf{X}\mathbf{I}_{\mathcal{S}}^T + \mathbf{G}_{\{\tilde{M}_i\}}\right)^{-1}\right),$$
$$\text{s.t.} \sum_{i=1}^{P} \log_2 \tilde{M}_i \leq \log_2 M,\ \tilde{M}_i \in \mathbb{Z}^+, i \in \mathcal{P},$$
$$\mathcal{S} = \{j | \tilde{M}_j > 1\}, |\mathcal{S}| \leq P. \tag{25}$$

To solve (25), we propose a greedy method starting with $\tilde{M}_i^{(0)} = 1$ for each $i \in \mathcal{N} \triangleq \{1, 2, \ldots, N\}$ and select a quantizer to assign an additional level, as in Algorithm 1. In the k-th iteration, we define the sampling set and bit allocation as $\mathcal{S}^{(k)}$ and $\{\tilde{M}_i^{(k)}\}$ respectively. The objective (25) is now

$$f(\{\tilde{M}_i^{(k)}\}) \triangleq \operatorname{Tr}\left(\mathbf{I}_{\mathcal{S}^{(k)}}\mathbf{H}\mathbf{I}_{\mathcal{S}^{(k)}}^T\left(\mathbf{I}_{\mathcal{S}^{(k)}}\mathbf{X}\mathbf{I}_{\mathcal{S}^{(k)}}^T + \mathbf{G}_{\{\tilde{M}_i^{(k)}\}}\right)^{-1}\right).$$

In each iteration, we select which $\tilde{M}_i$ to increment by approximating the derivative of $f(\cdot)$ w.r.t. each $\tilde{M}_i$. The derivative w.r.t. the integer $\tilde{M}_i$ is approximated by

$$\frac{\partial f(\{\tilde{M}_i^{(k)}\})}{\partial \tilde{M}_j} \approx q_j^{(k)} \triangleq f(\{\tilde{M}_i^{(k)} + \mathbf{1}_{i=j}\}) - f(\{\tilde{M}_i^{(k)}\}).$$

At the kth iteration, the approximated gradient vector $\mathbf{q}^{(k)}$ is

$$\nabla_{\{\tilde{M}_i^{(k)}\}} f(\{\tilde{M}_i^{(k)}\}) \approx \mathbf{q}^{(k)} \triangleq [q_1^{(k)}, q_2^{(k)}, \ldots, q_N^{(k)}]^T. \tag{26}$$

Our proposed greedy method iteratively updates $\{\tilde{M}_i\}$ while accounting for the constraints on the overall number of bits and the maximal number of actively qauntized samples imposed in (25). To that aim we define a set of possible sample indices $\mathcal{I}$, which is initialized to $\mathcal{N}$, i.e., all possible samples. In each iteration, we choose one quantizer index in $\mathcal{I}$ to assign an additional level, by selecting the one which we expect to contribute most substantially to the objective in (25), determined by the largest entry of (26). Since the number of actively quantized samples cannot go beyond P, once P different samples are actively quantized, we fix the set $\mathcal{I}$ to only consider the indexes of

Algorithm 2: Bit Allocation for a Given Graph Filter.

```
Input: Graph filter matrix $\mathbf{F}$, maximal number of samples $P$;
Output: Bit allocation $M_i$ and Sampling set $\mathcal{S}$;
Initialize: $k = 0$, bit allocation $\tilde{M}_i^{(0)} = 1$ for $\forall i \in \mathcal{N}$,
  possible indices $\mathcal{I} = \mathcal{N}$, and sampling set $\mathcal{S}^{(0)} = \emptyset$.
while $\sum_i \log_2 \tilde{M}_i^{(k)} \leq \log_2 M$ do
    Compute $q_i^{(k)}$ for each $i \in \mathcal{I}$;
    Update $\tilde{M}_i^{(k+1)} = \tilde{M}_i^{(k)} + \mathbf{1}_{i = \arg\min_{i \in \mathcal{I}} q_i^{(k)}}$;
    if $|\mathcal{S}^{(k)}| = P$ then
        Set possible indices $\mathcal{I} = \mathcal{S}^{(k+1)} = \mathcal{S}^{(k)}$;
    else
        Update $\mathcal{S}^{(k+1)} = \{i | \tilde{M}_i^{(k+1)} > 1\}$;
    end if
    $k = k + 1$;
end while
return $\{\tilde{M}_i\} = \{\tilde{M}_i^{(k-1)}\}$, $\mathcal{S} = \mathcal{S}^{(k-1)}$.
```

these samples subsequently. The overall process is summarized in Algorithm 2.

C. Alternating Optimization Algorithm Design

We proceed to solve (P3) for a given bit allocation $\{M_i\}$ and sampling set $\mathcal{S}$. In this case, (P3) is simplified to

$$\max_{F(\mathbf{\Lambda})} \operatorname{Tr}\left(\mathbf{I}_{\mathcal{S}} \mathbf{H} \mathbf{I}_{\mathcal{S}}^T \left(\mathbf{I}_{\mathcal{S}} \mathbf{X} \mathbf{I}_{\mathcal{S}}^T + \mathbf{G}\right)^{-1}\right). \tag{27}$$

The dependency of (27) on $F(\mathbf{\Lambda})$ is encapsulated in $\mathbf{H}$ and $\mathbf{X}$. Recalling the definition of the frequency domain graph filter in (Definition 3) and the imposed polynomial structure, i.e., $F(\mathbf{\Lambda}) = \sum_{i=0}^{K_0} \beta_i \mathbf{\Lambda}^i$, we aim to optimize (27) with respect to the coefficients $\{\beta_i\}$. Since $F(\mathbf{\Lambda})$ is diagonal, we do this by first optimizing its diagonal entries, denoted $\{\tilde{\lambda}_i\}_{i=1}^N$, after which we approximate them using a setof $\{\beta_i\}_{i=1}^{K_0}$.

To solve (27) with respect to $\{\tilde{\lambda}_i\}$, we consider the alternate optimization method. In particular, we optimize each $\tilde{\lambda}_i$ for fixed $\{\tilde{\lambda}_j\}_{j \neq i}$, repeating and updating this process for each i until convergence is achieved. Now, for a fixed index $i \in \mathcal{N}$, it follows from the definitions of $\mathbf{X}$ (see Lemma 2) and $\mathbf{G}$ in (10) that $\mathbf{I}_{\mathcal{S}} \mathbf{X} \mathbf{I}_{\mathcal{S}}^T + \mathbf{G} = \tilde{\lambda}_i^2 \mathbf{B} + \mathbf{C}$, where the $|\mathcal{S}| \times |\mathcal{S}|$ matrices $\mathbf{B}$ and $\mathbf{C}$ are independent of $\tilde{\lambda}_i$, and are defined as

$$(\mathbf{B})_{p,q} \triangleq m_{p,q} \left(\sigma_i^2 + \sigma^2\right) (\mathbf{U})_{(\mathcal{S})_p, i} (\mathbf{U})_{(\mathcal{S})_q, i},$$

$$(\mathbf{C})_{p,q} \triangleq m_{p,q} \sum_{j=1, j \neq i}^{N} \tilde{\lambda}_j^2 (\sigma_j^2 + \sigma^2) (\mathbf{U})_{(\mathcal{S})_p, j} (\mathbf{U})_{(\mathcal{S})_q, j},$$

with $m_{p,q} \triangleq 1 + \frac{2\eta^2}{3M_p^2} \mathbf{1}_{p=q}$. We can now write (and solve) the optimization of (27) w.r.t. a single $\tilde{\lambda}_i$ as

$$\max_{\tilde{\lambda}_i} \operatorname{Tr}\left(\mathbf{I}_{\mathcal{S}} \mathbf{H} \mathbf{I}_{\mathcal{S}}^T \left(\tilde{\lambda}_i \mathbf{B} + \mathbf{C}\right)^{-1}\right). \tag{28}$$

Algorithm 3: Constrained Graph Filter for Compression Design.

```
Input: Sampling set $\mathcal{S}$ and bit assignment $\{M_i\}$;
Output: Graph filter matrix $\mathbf{F}$;
Initialize: $\tilde{\lambda}_i^{(0)} = 1, i \in \mathcal{N}$.
for $k = 1, 2, \ldots, T$ do
    for $i \in \mathcal{S}$ do
        Update $\tilde{\lambda}_i^{(k)} = \tilde{\lambda}_i^*$ computed via Lemma 3;
    end for
    if $\left\|[\tilde{\lambda}_1^{(k)}, \ldots, \tilde{\lambda}_N^{(k)}] - [\tilde{\lambda}_1^{(k-1)}, \ldots, \tilde{\lambda}_N^{(k-1)}]\right\|_2^2 \leq \epsilon$
    then
        Break;
    end if
end for
Compute $F(\Lambda) = \mathrm{diag}\{\tilde{\lambda}_1^{(k-1)}, \tilde{\lambda}_2^{(k-1)}, \ldots, \tilde{\lambda}_N^{(k-1)}\}$;
return $\mathbf{F} = \mathbf{U} F(\mathbf{\Lambda}) \mathbf{U}^T$.
```

Lemma 3: The solution to (28) is $\tilde{\lambda}_i^* = \sqrt{\hat{\lambda}_i^*}$, where

$$\hat{\lambda}_i^* = \min_{\hat{\lambda}_i \geq 0} \sum_{j=1}^{P} \frac{a_j}{\hat{\lambda}_i + (\mathbf{\Lambda}_x)_{j,j}}. \tag{29}$$

In (29), we define $a_i \triangleq (\mathbf{A})_{i,i}^2 (\mathbf{\Lambda}_x)_{j,j} - \sum_{n=1, n \neq i}^{N} \hat{\lambda}_n (\mathbf{A})_{i,j}^2$, where $\mathbf{A} \triangleq \mathbf{U}_x^T \mathbf{B}^{-1/2} \mathbf{I}_{\mathcal{S}} \mathbf{U} \tilde{\mathbf{\Lambda}}$, with $\mathbf{U}_x$ and $\mathbf{\Lambda}_x$ being the eigenvectors and eigenvalues of $\mathbf{B}^{-1/2} \mathbf{C} \mathbf{B}^{-1/2}$, respectively.

Proof: The proof is given in Appendix F. ■

Problem (29) is a concave-convex fractional programming problem, which can be efficiently solved using the quadratic transform [33]. After $\tilde{\lambda}_i^*$ is obtained, we update the value of $F(\mathbf{\Lambda})$ and optimize the remaining $\tilde{\lambda}_i$. This procedure is repeated until a pre-specified convergence condition ϵ is met, or a maximal number of iterations T is reached. The resulting alternating algorithm for constrained graph filter design is summarized in Algorithm 3, where ϵ is the threshold and T is the maximum number of iterations.

Finally, the overall alternating optimization based algorithm for setting the frequency domain graph filter, sampling set, and bit allocation based on (P3) is summarized as Algorithm 4. The algorithm alternates between setting the frequency-domain graph filter for fixed sampling set and bit allocation via Algorithm 3, and optimizing the sampling set and bit allocation for fixed graph filter via Algorithm 2.

D. Discussion

Our proposed compression scheme is based on alternating optimization designing the sampling and quantization mappings to achieve a finite bit representation of the graph signal in a manner which allows its accurate reconstruction. This is achieved utilizing a sampling matrix which, instead of selecting a subset of the nodes as in [6], samples linear combinations of the graph signal elements using frequency domain graph filtering. The increased flexibility in the design of the sampling mapping is exploited to generate a sampling set such that the graph signal spectrum

Algorithm 4: Compression System Design with Frequency Domain Graph Filters.

```
Input: Graph Fourier basis matrix $\mathbf{U}^T$, maximal samples $P$;
Output: Graph filter $\mathbf{F}$, bit allocation $\{M_i\}$, sampling set $\mathcal{S}$;
Initialize: $\mathbf{F}^{(0)} = \mathbf{I}$, $k = 0$.
repeat
    Compute $\mathcal{S}^{(k)}$ and $\{M_i^{(k)}\}$ for $\mathbf{F}^{(k)}$ via Algorithm 2;
    Update $\mathbf{F}^{(k+1)}$ for $\mathcal{S}^{(k)}$ and $\{M_i\}^{(k)}$ via Algorithm 3;
    $k = k + 1$;
until $\|\mathbf{F}^{(k)} - \mathbf{F}^{(k-1)}\| \leq \epsilon$
return $\mathbf{F} = \mathbf{F}^{(k)}$, $\{M_i\} = \{M_i^{(k-1)}\}$, $\mathcal{S} = \mathcal{S}^{(k-1)}$.
```

can be accurately recovered from the quantized samples. In particular, we do this by accounting for the statistical moments of the noisy graph signal via the relaxed formulation (P3) when alternately tuning the sampling matrix and the quantization rule. We focus here on implementing such compression mechanisms using either unconstrained graph filters as well as structured frequency domain graph filters. However, one can consider additional forms of graph filtering which may be preferable in some applications, e.g., vertex domain graph filters [8]. Nonetheless, we leave the specialization of the proposed alternating optimization based sampling and quantization mechanism to alternative forms of graph filters for future work.

Finally, we note that our design considers a limit on the overall number of bits used by the quantizers to generate the finite bit representation. A less distorted finite bit representation can be obtained by replacing the uniform scalar quantizers with vector quantization [31], Ch. 23], though such mappings tend to be highly complex, as opposed to simplified uniform quantization. An alternative approach to improve upon the existing mechanism is to further compress $Q(\mathbf{\Psi x})$ via lossless source coding, e.g., entropy coding [34], Ch. 13], to $Q(\mathbf{\Psi x})$, as done in [35]. Doing so results in an equivalent representation whose number of bits is dictated by the entropy of $Q(\mathbf{\Psi x})$ (which is not larger than $\log_2 M$). This implies that the resulting finite bit representation can be further compressed, and motivates extending our analysis to constrain the entropy of $Q(\mathbf{\Psi x})$ rather than its bit budget, i.e., via entropy-constrained quantization [15], which we leave for future research.

V. Numerical Evaluations

In this section, we evaluate the performance of the proposed joint sampling and quantization methods for graph signal compression.[1] In Subsection V-A, we simulate graph signals representing measurements taken using parameters provided in GSP toolbox [36]. We then use the proposed scheme to compress graph signals representing real-world temperature data in Subsection V-B. Finally, in Subsection V-C, we apply the proposed sampling algorithm to image compression.

TABLE I
Simulation Parameters

Parameter	Symbol	Value
Number of nodes	N	100
Number of signals	N_0	1000
Bandwidth	K	20
Overall number of bits	$\log_2 M$	60
Number of quantizers	P	20
Power of noise	σ_0^2	-20dB

Throughput this section, we evaluate we use the MSE as the performance measure, defined as follows

$$\mathrm{MSE} = \mathbb{E}\{\|\hat{\mathbf{c}} - \mathbf{c}\|^2\}, \tag{30}$$

where $\hat{\mathbf{c}}$ and $\mathbf{c}$ are the recovered and the true graph signal frequency representations, respectively, and the expectation is computed via empirical averaging. In particular, we consider compressing the graph signal using the following schemes: (1) Joint sampling and quantization with unconstrained filters via Algorithm 1; (2) Separate sampling and quantization with non-identical quantizers as in [20]; (3) the MMSE estimate achievable of $\mathbf{U}_k\mathbf{c}$ which is achievable with infinite resolution quantization, i.e., without quantization constraints; (4) Joint sampling and quantization with identical quantizers as in [21]; and (5) Alternating sampling and quantization with frequency domain graph filters via Algorithm 4.

A. Synthetic Data

We begin with synthetic simulated scenario. Here, the graph signals represent measurements acquired by a sensor network comprised of $N = 100$ nodes, where each node is connected with its neighbors located within its transmission range, and the bandwidth is $K = 20$. We then take 1000 different noisy bandlimited graph signals with zero mean and covariance $\mathbf{C}_\mathbf{x}$, in which the non-zero GFT coefficients are randomized from $\mathcal{N}(0, (\Lambda^\dagger)_K)$, where Λ is the eigenvalue of $\mathbf{L}$. The specific simulation parameters are shown in Table 1.

We first numerically evaluate the MSE in compressing the graph signals versus the bit budget $\log_2 M$, which takes values in $[20, 120]$, while setting the Gaussian noise power to $\sigma_0^2 = -30$ dB. The results, depicted in Fig. 3. Observing Fig. 3, we note that Algorithm 1 achieves better performance than using separately designed sampling and quantization as well as applying the joint design with identical quantizers of [21]. The separate sampling and quantization mechanism of [20], in which the sampling operation is carried out by mere node selection without accounting for the underlying statistics of the signal, achieves the highest MSE values here, while the proposed schemes is within a small MSE gap of less than 0.02 from the MMSE for bit budgets as small as $\log_2 M = 40$. We also noted that Algorithm 4 allows achieving MSE which is comparable to

[1] The source code used in this section is available online in the following link: https://github.com/Pei65536/Graph_Signal_Compression.git

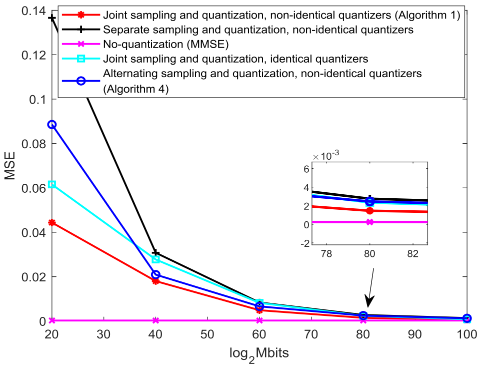


Fig. 3. MSE versus the number of bits for different schemes.

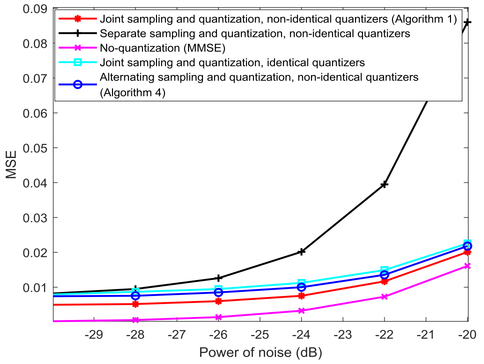
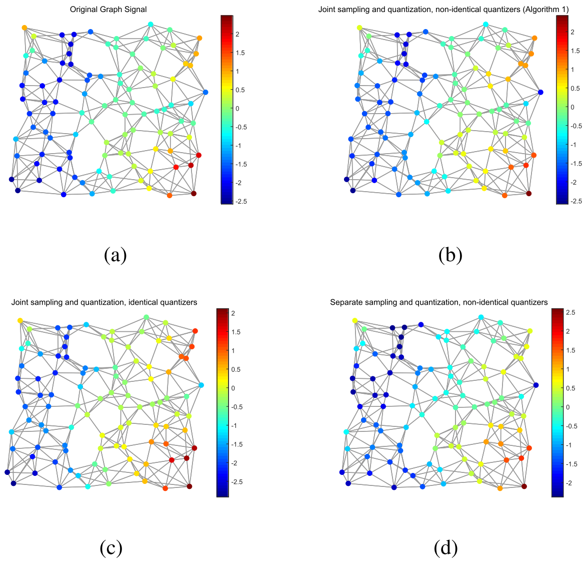


Fig. 4. MSE versus the variance of noise.

that achieved using unconstrained graph filters for bit budgets satisfying $\log_2 M \geq 40$.

Next, we numerically evaluate the effect of the additive noise power on the compression accuracy. To that aim, we compare the achievable MSEs of the schemes simulated in Fig. 3, while fixing the bit budget to be $\log_2 M = 60$, and letting the noise variance σ_0^2 grow from -30 dB to -20 dB. The results, depicted in Fig. 4, demonstrate that our proposed schemes maintains their improved MSE over the competing schemes for different levels of the noise variance. Our scheme is shown to be notably more robust to the presence of noise compared to all the reference techniques, except for the MMSE estimate, which is not constrained to any bit budget.

To visualize how the observed MSE gains are translated into a more faithful recovery of compressed graph signals, we depict a realization of a graph signal along with its reconstruction using scheme (1) and the benchmarks (2) and (4) in Fig. 5. Here, the noise level is set to $\sigma_0^2 = -30$dB, while the overall bit budget is $\log_2 M = 60$. Fig. 5 demonstrates that the proposed scheme based on joint sampling and quantization allows to compress the graph signal into a compact bit representation while resulting in recovered graph filters within a close match of the original graph signal. When using identical quantizers as in [21], the quality of the recovery signal in Fig. 5(d) is degraded and not as smooth as it is in Fig, 5(b). This follows since the low frequency components usually occupy a relatively large range, and using identical quantizers limits the recovery of the low frequency components. For the separate sampling and quantization scheme of [20], we observe in Fig. 5(c) poor recovery on some vertexes, which is caused by the amplification of quantization noise and additive Gaussian noise in restoring unsampled vertexes.


Original Graph Signal
Joint sampling and quantization, non-identical quantizers (Algorithm 1)
Joint sampling and quantization, identical quantizers
Separate sampling and quantization, non-identical quantizers
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5

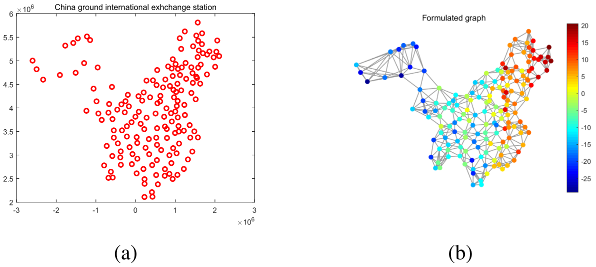
(a)

(b)

(c)

(d)

Fig. 5. Synthetic graph signal recovery experiments: (a) Graph signal; (b) Reconstructed signal via Algorithm 1; (c) Reconstructed signal with separate sampling and quantization with non-identical quantizers; (d) Reconstructed signal with joint sampling and quantization with identical quantizers.



(a)

(b)

Fig. 6. (a) Geographical location of China ground sensors. (b) A temperature example in formulated graph structure.

B. Non-Synthetic Graph Signals

In order to further illustrate the practicability of our graph signal compression algorithms, we apply the proposed Algorithm 1 and Algorithm 4 to compress non-synthetic graph signals representing temperature data.[2] The data is obtained from the China Meteorological Data Center, representing measurements taken in January of each year from 2018 to 2021 via China's meteorological sensors whose distribution is is shown in Fig. 6(a). The location of each sensor is determined by its latitude and longitude. It is noted that temperature data distribution depends not only on the geographical location, but is also affected by the surrounding environment and human actions. Therefore, we

[2] The dataset could be downloaded in the following link: http://data.cma.cn

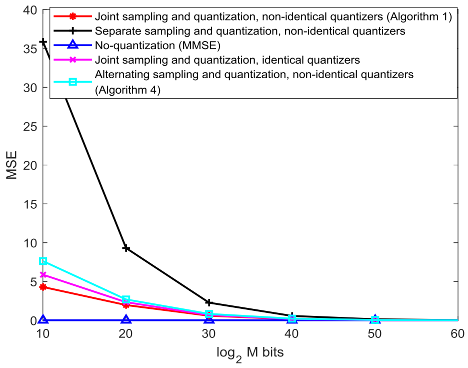


Fig. 7. MSE versus the number of bits.

construct the graph structure for these measurements based on the smoothness of the data, rather than directly constructing the graph structure based on the distance threshold, e.g., as in K-nearest neighbours construction. We formulate an optimization problem with setting the smoothness as the objective. The specific processing has been investigate in existing works [12]. It can learn graph structure via input data, which results that the graph signals with smooth variations on the resulting Laplacian matrix. An example of a resulting graph signal is illustrated in Fig. 6(b). Note that the temperature data is selected in a short time period of each year. Specifically, the temperature in a fixed node of graph (a sensor in reality) tends to a stable state, which means that we can impose a Gaussian probabilistic prior on the frequency presentation of these graph signals and the latent noise. The statistical information of the temperature data shows that when the bandwidth is greater than 10, the frequency domain component is close to 0. Therefore, we assume that the bandwidth of the data is $K = 10$.

The achievable reconstruction MSE for these non-synthetic graph signals versus the overall number of bits is depicted in Fig. 7. It is noted that in this scenario there is no noise signal that is manually introduced, and the equivalent noise is that which stems from the error in approximating the graph signal as have bandwidth of $K = 10$. We observe in Fig. 7 that the MSE trends versus the total number of bits is preserved as in the synthetic case reported in the previous subsection. Specifically, it is found that compression with unconstrained graph filters via Algorithm 1 outperforms all considered bit-constrained benchmarks, and that similar performance is achieved with frequency domain graph filters for at least 20 bits using Algorithm 4.

C. Application in Image Compression

GSP is not only applied to the data with irregular structures, but can also be used to represent signals such as images, video signals, etc. Here, we apply our proposed Algorithm 1, which designs joint sampling and quantization using unconstrained graph filters to image compression. Due to the self-similarity in images, the same or similar structures are likely to recur throughout. For simplicity, we apply regular line and grid graph topologies, but with unequal edge weights that can adapt to the specific characteristic of an image or a set of images [37].

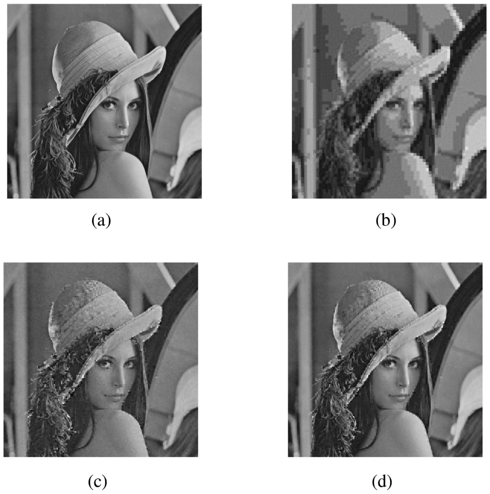

(a) (b) (c) (d)

Fig. 8. Image compression. (a) The original image; (b) The compressed image via DCT; (c) The compressed image via joint sampling and quantization with identical quantizers; (d) The compressed image via Algorithm 1.

In particular, we use the classic ‘Lena‘ image comprised 512×512 pixels, which we model as a graph signal with a grid topology with edge weights obtained by low-resolution GFT [37]. The image is thus expressed as a concatenation of 16384 signals, each representing 4×4 patches. For each graph signal, we approximate its bandwidth to be $K = 4$, i.e., we discard high-frequency components, which is relatively quite small. The statistical moments used in our derivation are obtained by averaging over the patches of the image. We compare the performance of three image compression schemes: 1) The Discrete Cosine Transform (DCT); 2) Separate sampling with quantization with non-identical quantizers [20]; 3) The proposed Algorithm 1. For consistency, we set the block size of DCT to be 4×4. We set $\log_2 M = 64$ for each GFT block and DCT block. The results, depicted in Fig. 8, indicate that the proposed mechanism for joint sampling and quantization which considers graph signals can also be beneficial for image compression, resulting in a faithful recovery of natural images compressed into a low bit representation.

VI. Conclusion

In this work, we studied the compression of bandlimited graph signals into a finite-length sequence of bits by joint sampling and quantization. Our derivation is based on the identified similarity between graph signal compression and task-based quantization, which is a framework for configuring ADCs with analog pre-processing. We formulated the design problem, which is in general shown to be non-convex. We then presented a relaxed formulation considering linear recovery with non-overloaded quantizers. The relaxed problem was used to design sampling mechanisms for a given allocation of the available bit budget among the quantizers, which in turn is used to derive a greedy bit allocation algorithm. Next, we proposed a sampling operator as a row-selection matrix with a graph filter, restricted to combining elements corresponding to the neighbouring nodes of the graph.

We applied the proposed algorithm to both synthetic and non-synthetic data. Numerical results show that the proposed scheme compresses high dimensional graph signals into a limited amount of bits while allowing their recovery with an error within a small gap from the MMSE, achievable with infinite resolution quantziation.

APPENDIX

A. *Proof of Proposition 1*

We first note that the MSE achievable using the linear recovery matrix $\mathbf{\Phi}^*$ in Lemma 1 is given by

$$\begin{aligned}\mathrm{MSE}(\mathbf{\Psi}) &= \mathrm{Tr}\left(\mathbf{\Gamma}^*\mathbf{C}_{\mathbf{x}}\mathbf{\Gamma}^{*T}\right)\\ &\quad - \mathrm{Tr}\left(\mathbf{\Gamma}^*\mathbf{C}_{\mathbf{x}}\mathbf{\Psi}^T\left(\mathbf{\Psi}\mathbf{C}_{\mathbf{x}}\mathbf{\Psi}^T+\mathbf{G}\right)^{-1}\mathbf{\Psi}\mathbf{C}_{\mathbf{x}}\mathbf{\Gamma}^{*T}\right).\end{aligned} \tag{A.1}$$

By discarding the constant term $\mathrm{Tr}(\mathbf{\Gamma}^*\mathbf{C}_{\mathbf{x}}\mathbf{\Gamma}^{*T})$ in (A.1), the matrix $\mathbf{\Psi}^*$ that minimizes the MSE can be obtained via

$$\hat{\mathbf{\Psi}}^* = \arg\max_{\hat{\mathbf{\Psi}}}\mathrm{Tr}\big(\hat{\mathbf{\Gamma}}\hat{\mathbf{\Psi}}^T\big(\hat{\mathbf{\Psi}}\hat{\mathbf{\Psi}}^T+\mathbf{I}\big)^{-1}\hat{\mathbf{\Psi}}\hat{\mathbf{\Gamma}}^T\big), \tag{A.2}$$

where $\hat{\mathbf{\Psi}} \triangleq\!= \mathbf{G}^{-1/2}\mathbf{\Psi}\mathbf{C}_{\mathbf{x}}^{1/2}, \hat{\mathbf{\Gamma}} \triangleq\!= \mathbf{\Gamma}^*\mathbf{C}_{\mathbf{x}}^{1/2}$. Now, by writing the singular value decomposition $\hat{\mathbf{\Psi}} = \mathbf{U}_\Psi\mathbf{\Xi}_\Psi\mathbf{V}_\Psi^T$, and recalling the definition of $\mathbf{G}$, we obtain that $\frac{\mathbf{G}_{i,i}M_i^2}{\eta^2} = (\mathbf{\Psi}\mathbf{C}_{\mathbf{x}}\mathbf{\Psi}^T)_{i,i} = (\mathbf{G}^{1/2}\hat{\mathbf{\Psi}}\hat{\mathbf{\Psi}}^T\mathbf{G}^{1/2})_{i,i}$, which results in

$$(\mathbf{U}_\Psi\mathbf{\Xi}_\Psi^2{\mathbf{U}_\Psi}^T)_{i,i} = \frac{M_i^2}{\eta^2}. \tag{A.3}$$

According to majorization theory [38], (A.3) is satisfied if and only if $\{M_i^2/\eta^2\} \prec \{\alpha_i\}$, where $\alpha_i = (\mathbf{\Xi}_\Psi)_{i,i}^2$. Then problem (A.2) is transformed into

$$\begin{aligned}&\max_{\mathbf{V}_\Psi,\mathbf{\Xi}_\Psi}\ \mathrm{Tr}\left(\mathbf{V}_\Psi^T\hat{\mathbf{\Gamma}}^T\hat{\mathbf{\Gamma}}\mathbf{V}_\Psi\mathbf{\Xi}_\Psi^T\big(\mathbf{\Xi}_\Psi\mathbf{\Xi}_\Psi^T+\mathbf{I}\big)^{-1}\mathbf{\Xi}_\Psi\right),\\ &\qquad \text{s.t.}\ \left\{\frac{3M_i^2}{2\eta^2}\right\} \prec \{\alpha_i\}, \alpha_i = (\mathbf{\Xi}_\Psi)_{i,i}^2.\end{aligned} \tag{A.4}$$

We note that $\mathbf{\Xi}_\Psi^T(\mathbf{\Xi}_\Psi\mathbf{\Xi}_\Psi^T+\mathbf{I})^{-1}\mathbf{\Xi}_\Psi$ is a diagonal matrix with diagonal elements $\frac{\alpha_i}{\alpha_i+1}, i \in \mathcal{P}$, which are a descending sequence since $\{\alpha_i\}$ is arranged in descending order. Then, the optimal $\mathbf{V}_\Psi$ should be the right singular vectors matrix of $\hat{\mathbf{\Gamma}}$. By substituting (2) and (4) into the definition of $\hat{\mathbf{\Gamma}}$, we obtain $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Lambda}}(\tilde{\mathbf{\Lambda}}^{-1})_K\mathbf{U}^T$. Note that $\hat{\mathbf{\Lambda}}(\tilde{\mathbf{\Lambda}}^{-1/2})_K$ is a diagonal matrix with descending diagonal elements, i.e., $\mathbf{V}_\Psi^T = \mathbf{U}^T$. Consequently, $\tilde{\mathbf{\Gamma}} = \mathbf{V}_\Psi^T\hat{\mathbf{\Gamma}}^T\hat{\mathbf{\Gamma}}\mathbf{V}_\Psi$ is a diagonal matrix, whose diagonal elements arrange in a descending order. The remaining optimization of (A.4) is thus

$$\begin{aligned}&\max_{\alpha_i}\sum_{i=1}^{P}\frac{\lambda_{\mathbf{\Gamma},i}^2\alpha_i}{\alpha_i+1},\\ &\text{s.t.}\sum_{i=1}^{P}\alpha_i = \sum_{i=1}^{P}\frac{M_i^2}{\eta^2}, \sum_{i=1}^{p}\alpha_i \geq \sum_{i=1}^{p}\frac{M_i^2}{\eta^2}, \forall p \in \mathcal{P},\end{aligned} \tag{A.5}$$

where $\lambda_{\mathbf{\Gamma},i}$ is the i-th singular value of $\hat{\mathbf{\Gamma}}$. The constraints in (A.5) are the expanded form of $\left\{\frac{M_i^2}{\eta^2}\right\} \prec^w \{\alpha_i\}$. We note that (A.5) is convex. Once $\{\alpha_i\}$ are obtained, the resulting sampling matrix is $\mathbf{\Psi}^* = \mathbf{G}^{1/2}\mathbf{U}_\Psi\mathbf{\Xi}_\Psi\mathbf{V}_\Psi^T{\mathbf{C}_{\mathbf{x}}}^{-1/2}$. While the entries of $\mathbf{G}$ depend on $\mathbf{\Psi}$, we can still obtain the sampling matrix by letting $g_i = \mathbf{G}_{i,i}^{1/2}$, and substituting the expression of $\mathbf{\Psi}^*$ into the definition of $\mathbf{G}$. This results in $\frac{\eta^2(\mathbf{\Psi}\mathbf{C}_{\mathbf{x}}\mathbf{\Psi}^T)_{i,i}}{M_i^2} = g_i^2$, and thus $(\mathbf{G}^{1/2}\mathbf{U}_\Psi\mathbf{\Xi}_\Psi^2{\mathbf{U}_\Psi}^T\mathbf{G}^{1/2})_{i,i} = \frac{g_i^2M_i^2}{\eta^2}$, which holds for any g_i by (A.3). We can thus compute $\mathbf{\Psi}$ with $\mathbf{G} = \mathbf{I}$, obtaining $\mathbf{\Psi}^* = \mathbf{U}_\Psi\mathbf{\Xi}_\Psi\tilde{\mathbf{\Lambda}}^{-1/2}\mathbf{U}^T$, concluding the proof. ■

B. *Proof of Theorem 1*

Lemma 1 characterizes the MSE minimizing sampling operator $\mathbf{\Psi}$ for fixed bit allocation $\{M_i\}$. To optimize $\{M_i\}$, we focus on the case where $\{M_i\}$ are not limited to be integer, resulting in the following formulation

$$\begin{aligned}&\max_{\{M_i\},\{\alpha_i\}}\sum_{i=1}^{P}\frac{\lambda_{\mathbf{\Gamma},i}^2\alpha_i}{\alpha_i+1},\\ &\qquad\text{s.t.}\left\{\frac{M_i^2}{\eta^2}\right\} \prec \{\alpha_i\}, \sum_{i=1}^{P}\log_2 M_i \leq \log_2 M,\ M_i \geq 1.\end{aligned} \tag{B.1}$$

Problem (B.1) is non-convex and difficult to solve, Hence, we first propose the following lemma to simplify (B.1).

Lemma B.1: The solution of (B.1) satisfies $\frac{M_i^2}{\eta^2}=\alpha_i, \forall i \in \mathcal{P}$.

Proof: We first recall the following result from [38]. Let $\phi : \mathcal{D}_n \to \mathbb{R}$ be a real-valued function continuous on $\mathcal{D}_n \triangleq \{\mathbf{x} \in \mathbb{R}^n : x_1 \geq \ldots \geq x_n\}$ and continuously differentiable on the interior of $\mathcal{D}_n$. Then ϕ is Schur-convex (Schur-concave) on $\mathcal{D}_n$ if and only if $\frac{\partial\phi(\mathbf{x})}{\partial x_i}$ is decreasing (increasing) in $i = 1, \ldots, n$.

Assume that $\{\acute{M}_i\}$ is the optimal bit allocation and $\{\acute{\alpha}_i\}$ is the corresponding diagonal elements satisfying $\{\acute{M}_i^2/\eta^2\} \prec \{\acute{\alpha}_i\}$. Denote $M^{(1)} = \sum_{i=1}^{P}\log_2\acute{M}_i^2/\eta^2$ and $M^{(2)} = \sum_{i=1}^{P}\acute{\alpha}_i$. Since $\sum\log_2 M_i$ is Schur-concave with respect to M_i, then $M^{(1)} \geq M^{(2)}$. Thus, one can propose another $\{\grave{M}_i\}$ and $\{\grave{\alpha}_i\}$ satisfying $\grave{M}_i^2/\eta^2 = \grave{\alpha}_i = \acute{\alpha}_i^{\kappa_0}$, where $\kappa_0 = M^{(1)}/M^{(2)} \geq 1$. Clearly, $\{\grave{M}_i\}, \{\grave{\alpha}_i\}$ are feasible for (B.1), with better MSE than $\{\acute{\alpha}_i\}$, proving the lemma. ■

By Proposition 1, the sampling matrix should be

$$\mathbf{\Psi}^* = \mathbf{U}_\Psi\mathbf{\Xi}_\Psi\tilde{\mathbf{\Lambda}}^{-1/2}\mathbf{U}^T. \tag{B.2}$$

Substituting Lemma B.1 into (B.2), the matrix $\mathbf{U}_\Psi$ becomes the identity. Hence, the sampling matrix would be a diagonal matrix multiplied by the unitary $\mathbf{U}^T$. The proof of Proposition 1 reveals that the diagonal matrix $\mathbf{G}^{-1/2}$ which is leftmost in the expression for $\mathbf{\Psi}^*$ could be arbitrarily set. As a result, the MSE is minimized by the sampling matrix $\mathbf{\Psi}^* = \mathbf{U}_K^T$.

To determine $\{M_i\}$, we write problem (B.1) as $\max_{\{M_i\}}\sum_{i=1}^{P}\frac{\lambda_{\mathbf{\Gamma},i}^2M_i^2}{M_i^2+\eta^2}$, s.t. $\sum_{i=1}^{P}\log_2 M_i \leq \log_2 M$, $M_i \geq 1$, $M_i \geq M_{i+1}$, which is still non-convex. To tackle this, we obtain a local solution and prove its optimality. According to Lagrange Duality theorem,we introduce the multipliers $\beta^* \in \mathbb{R}$ and $\nu^* \in \mathbb{R}^P$. While this procedure does not impose monotonicity on $\{M_i\}$, it is implicitly maintained since

$\lambda_{\mathbf{\Gamma},i}$ are arranged in a descending order. For simplicity, we denote $m_i = M_i^2$, and write the KKT constraints as $\sum_{i=1}^{P} \log_2 m_i^* \leq 2\log_2 M$, with $m_i^* \geq 1$, $\nu_i^*(m_i^* - 1) = 0$ and $\frac{\lambda_{\mathbf{\Gamma},i}^2 \eta^2}{(m_i+\eta^2)^2} - \nu_i^* - \frac{\beta^*}{m_i} = 0$, for each $i \in \mathcal{P}$. Eliminating the slack variable ν_i^*, we obtain that $\left(\frac{\beta^*}{m_i} - \frac{\lambda_{\mathbf{\Gamma},i}^2 \eta^2}{(m_i+\eta^2)^2}\right)(m_i^* - 1) = 0$ should hold.

We focus on the equation $\frac{\beta^*}{m_i} - \frac{\lambda_{\mathbf{\Gamma},i}^2 \eta^2}{(m_i+\eta^2)^2} = 0$, which is solvable if and only if $\beta^* \leq \lambda_{\mathbf{\Gamma},i}^2/4$. The local optimal m_i is obtained here by (18). The optimal bit allocation is expressed by the dual coefficient β^*, which satisfies the sum bits constraint $f(\beta) = \sum_{i=1}^{P} \log_2 m_i^* \leq 2\log_2 M$. Similar as the proof of lemma B.1, we obtain that $f(\beta) = 2\log_2 M$ should be satisfied. Further, $f(\beta)$ is a decreasing function with respect to β, i.e., the optimal β^* is the only one. ■

C. *Proof of Corollary 1*

We first focus on the necessary condition. Due to the definition of M_a, $(M_a)^P := (\lfloor M^{1/P} \rfloor)^P \leq M$, and $M \geq M_a^P$ should be obviously satisfied. We then assume that $M \geq M_a^P + M_a^{P-1}$, there exists another scheme of bit assignment as $[M_a + 1, M_a, \ldots, M_a]$ obtained by Algorithm 1, which against the previous assumption. The range of M is thus determined, which corresponds to the necessity of (21a).

We note that $\lambda_{\tilde{\Gamma},i}$ is a descending sequence, thus $g_i(M_i) \geq g_j(M_j)$ for $M_i = M_j$, and $g_i(M_i) > g_i(M_i + 1)$ for $M_i \geq 1$. When $M_i = M_j = M_a$ for each $i \in \mathcal{P}$, we define $k_0 = (M_a)^P - 1$, as we discussed above, the state $M_i^{(k_0)}$ should be $[M_a, M_a, \ldots, M_a]$ and $(k_0 + 1)$-th iteration should select P-th quantizer rather than others, which means that $g_P(M_a - 1) > g_i(M_a)$ for each $i \in \mathcal{P}$, and then simplified as $g_P(M_a - 1) > g_1(M_a)$. The necessity of (21b) is thus proven.

For the sufficient case, we assume both (21a) and (21b) are satisfied. Greedy-based method is applied in Algorithm 1, we try to describe the bit assignment process. In k-th iteration, the particular index is selected by $e = \arg\min_i g_i(M_i^{(k)})$. For example, we assume that in k_1-th iteration, the state of $M_1^{(k_1-1)}$ change from $M_a - 1$ to M_a, i.e., $M_1^{(k_1)} = M_a$. Here the other bit assignment should satisfy that $M_i(k_1) < M_a$ for each $i = 2, \ldots, P$ since $g_i(M_i) \geq g_j(M_j)$ for $M_i = M_j$. When (21b) is satisfied, it holds that $g_1(M_a) < g_P(M_a - 1) < g_{P-1}(M_a - 1) \cdots < g_2(M_a - 1)$. This, before the state of M_P change from $M_a - 1$ to M_a, the state of M_1 would not be changed. Further, we assume that in k_P-th iteration, the state of $M_P^{(k_P-1)}$ change from $M_a - 1$ to M_a, i.e., $M_P^{(k_P)} = M_a$. Consequently, $g_{P-1}(M_a) < g_{P-2}(M_a) \cdots < g_1(M_a) < g_P(M_a - 1)$ As a result, after the k_P-th iteration, $M_i = M_a$ for each $i = 1, \ldots, P$. Since (21a) limit the upper bound of M, the greedy-based algorithm would be ended here due to the termination condition, concluding the proof. ■

D. *Proof of Corollary 2*

We first focus on the case when sum of bits is limited, which may lead to $P < K$. We would complete this proof in two parts, in which we separately propose that

$$P < i \quad \text{if} \quad \log_2 M < l_i + 1, \qquad \text{(PartI)}$$

$$P \geq i \quad \text{if} \quad \log_2 M \geq l_i + 1. \qquad \text{(PartII)}$$

Define P^* as the number of quantizers obtained by Algorithm 1, and its bit allocation as $\{M_1^*, M_2^*, \ldots, M_{P^*}^*\}$. For any $P \leq P^*$, we assume that P-th quantizer is added in the k_0-th iteration, i.e., $P = \arg\max_{k_0} \mathbf{g}^{(k_0)}$, thus

$$g_P(1) \geq g_j(M_j^{(k_0)}) \geq g_j(M_j^*), \forall j < P. \qquad \text{(D.1)}$$

For the proof of the (PartI), we assume that $P \geq i$ and $\log_2 M < l_i + 1$. Then, we define the bit allocation as $\{\dot{M}_1, \dot{M}_2, \ldots, \dot{M}_P\}$, which should satisfy $g_i(\dot{M}_i) \leq g_P(1)$. It follows that $\dot{M}_i \geq \lceil \tilde{l}_j \rceil$. Then we obtain $\sum_{i=1}^{P} \dot{M}_i \geq \sum_{i=1}^{P-1} \dot{M}_i + 1 \geq l_i + 1$, which contradicts the assumption.

To prove (PartII), consider the k_1th iteration, where $g_i(M_i^{(k)}) \geq g_P(1)$ and $g_i(M_i^{(k)} + 1) \leq g_P(1)$ for $i < P$. Note that $M_i^{(k)} \leq \tilde{l}_j \leq M_i^{(k)} + 1$ and $\lceil \tilde{l}_j \rceil = M_i^{(k)} + 1$. Thus, for the $(k_1 + P - n)$th iteration, where $1 \leq n \leq P - 1$, $P = \arg\max_k \mathbf{g}^{(k)}$ is obtained, proving (PartII). Further, when $\log_2 M \geq l_K$, the Kth quantizer is added, and there is no $(K+1)$th quantizer since the gradient vector satisfies $g_i(M_i) = 0$ for $i > K$. This results in (22). ■

E. *Proof of Lemma 2*

To prove the equivalence between (P2) and (P3), we first note that that Lemma 1, which characterizes the MSE minimizing digital recovery filter, holds for any sampling matrix $\mathbf{\Psi}$, including those representing frequency-domain graph filtering. Consequently, by substituting the resulting recovery matrix $\mathbf{\Phi} = \mathbf{\Gamma}^* \mathbf{C_x} \mathbf{\Psi}^T (\mathbf{\Psi} \mathbf{C_x} \mathbf{\Psi}^T + \mathbf{G})^{-1}$, we arrive at $\max_{\mathbf{\Psi}, \{M_i\}} \mathrm{Tr}(\mathbf{\Gamma}^* \mathbf{C_x} \mathbf{\Psi}^T (\mathbf{\Psi} \mathbf{C_x} \mathbf{\Psi}^T + \mathbf{G})^{-1} \mathbf{\Psi} \mathbf{C_x} \mathbf{\Gamma}^{*T})$, s.t. $\sum_{i=1}^{P} \log_2 M_i \leq \log_2 M, M_i \in \mathbb{Z}^+$. By setting $\mathbf{\Psi} = \mathbf{I}_{\mathcal{S}} \mathbf{U}_K F(\mathbf{\Lambda}) \mathbf{U}_K^T$ to represent frequency-domain filtering, we transform the objective into (P3), proving the lemma. ■

F. *Proof of Lemma 3*

We first transform $(\tilde{\lambda}_i^2 \mathbf{B} + \mathbf{C})^{-1}$ into

$$\begin{aligned} (\tilde{\lambda}_i^2 \mathbf{B} + \mathbf{C})^{-1} &= \mathbf{B}^{-1/2} \left(\tilde{\lambda}_i^2 \mathbf{I} + \mathbf{B}^{-1/2} \mathbf{C} \mathbf{B}^{-1/2}\right)^{-1} \mathbf{B}^{-1/2} \\ &\overset{(a)}{=} \mathbf{B}^{-1/2} \left(\tilde{\lambda}_i^2 \mathbf{I} + \mathbf{U}_x \lambda_x \mathbf{U}_x^T\right)^{-1} \mathbf{B}^{-1/2} \\ &= \mathbf{B}^{-1/2} \mathbf{U}_x \left(\tilde{\lambda}_i^2 \mathbf{I} + \mathbf{\Lambda}_x\right)^{-1} \mathbf{U}_x^T \mathbf{B}^{-1/2}, \qquad \text{(F.1)} \end{aligned}$$

where (a) follows from the eigenvalue decomposition, where $\mathbf{U}_x$ is a unitary matrix since both $\mathbf{B}$ and $\mathbf{C}$ are symmetric. Substituting (F.1) and the definition of $\mathbf{A}$ into (28) yields

$$\max_{\tilde{\lambda}_i} \mathrm{Tr}(\mathbf{A} \tilde{\mathbf{\Lambda}}^2 F(\mathbf{\Lambda})^2 \mathbf{A}^T (\tilde{\lambda}_i^2 \mathbf{I} + \mathbf{\Lambda}_x)^{-1}). \qquad \text{(F.2)}$$

Noting that $\tilde{\lambda}_i$ is in form of $\tilde{\lambda}_i^2$ in (F.2), thus we define $\hat{\lambda}_i = \tilde{\lambda}_i^2$, simplifying (F.2) into: $\max_{\hat{\lambda}_i \geq 0} \sum_{j=1}^{P} \frac{\sum_{n=1}^{N} \hat{\lambda}_n (\mathbf{A})_{j,n}^2}{\hat{\lambda}_i + (\mathbf{\Lambda}_x)_{j,j}}$. Substituting the definition of $\{a_i\}$ yields (29). ■

References

[1] P. Li, N. Shlezinger, H. Zhang, B. Wang, and Y. C. Eldar, "Graph signal compression via task-based quantization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 5514–5518.

[2] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[3] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[4] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[5] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[6] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[7] Y. Tanaka and Y. C. Eldar, "Generalized sampling on graphs with subspace and smoothness priors," *IEEE Trans. Signal Process.*, vol. 68, pp. 2272–2286, 2020.

[8] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling on graphs: From theory to applications," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 14–30, Nov. 2020.

[9] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[10] A. C. Yağan and M. T. Özgen, "Spectral graph based vertex-frequency wiener filtering for image and graph signal denoising," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 226–240, 2020.

[11] A. Heimowitz and Y. C. Eldar, "A Markov variation approach to smooth graph signal interpolation," 2018, *arXiv:1806.03174*.

[12] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 2016.

[13] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2017, pp. 6508–6512.

[14] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 34–47, Jan. 2018.

[15] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[16] L. F. O. Chamon and A. Ribeiro, "Finite-precision effects on graph filters," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2017, pp. 603–607.

[17] I. C. M. Nobre and P. Frossard, "Optimized quantization in distributed graph signal processing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 5376–5380.

[18] L. B. Saad, E. Isufi, and B. Beferull-Lozano, "Graph filtering with quantization over random time-varying graphs," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2019, pp. 1–5.

[19] P. Di Lorenzo, S. Barbarossa, and P. Banelli, "Optimal power and bit allocation for graph signal interpolation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 4649–4653.

[20] Y. H. Kim and A. Ortega, "Toward optimal rate allocation to sampling sets for bandlimited graph signals," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1364–1368, Sep. 2019.

[21] N. Shlezinger, Y. C. Eldar, and M. R. D. Rodrigues, "Hardware-limited task-based quantization," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5223–5238, Oct. 2019.

[22] N. Shlezinger, Y. C. Eldar, and M. R. Rodrigues, "Asymptotic task-based quantization with application to massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3995–4012, Aug. 2019.

[23] S. Salamtian, N. Shlezinger, Y. C. Eldar, and M. Medard, "Task-based quantization for recovering quadratic functions using principal inertia components," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 390–394.

[24] P. Neuhaus, N. Shlezinger, M. Dörpinghaus, Y. C. Eldar, and G. Fettweis, "Task-based analog-to-digital converters," *IEEE Trans. Signal Process.*, vol. 69, pp. 5403–5418, 2021.

[25] N. Shlezinger and Y. C. Eldar, "Task-based quantization with application to MIMO receivers," *Commun. Inf. Syst.*, vol. 20, pp. 131–162, 2020.

[26] N. Shlezinger and Y. C. Eldar, "Deep task-based quantization," *Entropy*, vol. 23, no. 1, 2021, Art. no. 104.

[27] F. Gama, E. Isufi, G. Leus, and A. Ribeiro, "Graphs, convolutions, and neural networks: From graph filters to graph neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 128–138, Nov. 2020.

[28] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[29] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[30] B. Widrow, I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.

[31] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," *Lecture Notes for ECE563*, Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 2014, pp. 2012–2017.

[32] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design Via Majorization Theory*. Boston, MA, USA: Now Publishers, 2007.

[33] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[35] N. Shlezinger, M. Chen, Y. C. Eldar, H. V. Poor, and S. Cui, "UVeQFed: Universal vector quantization for federated learning," *IEEE Trans. Signal Process.*, vol. 69, pp. 500–514, 2021.

[36] N. Perraudin et al., "GSPBOX: A toolbox for signal processing on graphs," 2014, *arXiv:1408.5781*.

[37] W. Hu, G. Cheung, A. Ortega, and O. C. Au, "Multiresolution graph Fourier transform for compression of piecewise smooth images," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 419–433, Jan. 2015.

[38] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, vol. 143, Berlin, Germany: Springer, 1979.

Pei Li received the B.E. degree from Qingdao University, Qingdao, China. He is currently working toward the Ph.D. degree with the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include graph signal processing and graph neural network.

Nir Shlezinger (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University, Be'er Sheva, Israel, in 2011, 2013, and 2017, respectively. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Ben-Gurion University. From 2017 to 2019, he was a Postdoctoral Researcher with the Technion and with the Weizmann Institute of Science, Rehovot, Israel, from 2019 to 2020. His research interests include communications, information theory, signal processing, and machine learning. He was the recipient of the FGS prize for outstanding research achievements, from the Weizmann Institute of Science.

Haiyang Zhang (Member, IEEE) received the B.S. degree in communication engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2009, the M.S. degree in information and communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2012, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2017. He is currently an Assistant Professor with the School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications. From 2017 to 2020, he was a Postdoctoral Research Fellow with the Singapore University of Technology and Design, Singapore, and with the Weizmann Institute of Science, Rehovot, Israel, from 2020 to 2022. His research interests include 6G near-field communications, learning and sampling theory, and physical-layer security.

Baoyun Wang (Member, IEEE) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997. From 1999 to 2000, he was a Postdoctoral Research Fellow with the Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang, South Korea, with the Department of Electronic Engineering, City University of Hong kong, Hong Kong, from 2000 to 2002, and the Department of Mathmatics and Computer Science, University of Sydney, Camperdown, NSW, Australia, from 2004 to 2005. He is currently a Full Professor in electrical engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. His research intreasts include information theory, statistical signal processing, graph theory, and their applicaions in wireless communications.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics in 1995 and the B.Sc. degree in electrical engineering in 1996 from Tel-Aviv University, Tel-Aviv, Israel, and the Ph.D. degree in electrical engineering and computer science in 2002 from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in engineering, and Visiting Professor with Stanford. She is also a Visiting Professor with MIT, Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging and optics. Dr. Eldar was the recipient of the many Awards for Excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award in 2013, IEEE/AESS Fred Nathanson Memorial Radar Award in 2014, IEEE Kiyo Tomiyasu Award in 2016, Michael Bruno Memorial Award from the Rothschild Foundation, Weizmann Prize for Exact Sciences, Wolf Foundation Krill Prize for Excellence in Scientific Research, Henry Taub Prize for Excellence in Research (twice), Hershel Rich Innovation Award (three times), Award for Women with Distinguished Contributions, Andre and Bella Meyer Lectureship, Career Development Chair at the Technion, Muriel & David Jacknow Award for Excellence in Teaching, and Technion's Award for Excellence in Teaching (two times). She is also the recipient of several best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, was selected as one of the 50 most influential women in Israel and in Asia, and is a highly cited Researcher. She is a Member of the Israel Academy of Sciences and Humanities (elected 2017), EURASIP Fellow, Horev Fellow of the Leaders in Science and Technology Program with the Technion, an Alon Fellow, and a Member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a Member of the IEEE Sensor Array and Multichannel Technical Committee and with several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, Member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing Technical Committees, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was a Co-chair and Technical Co-chair of several international conferences and workshops. She is author of the book *Sampling Theory: Beyond Bandlimited Systems* and coauthor of five other books published by Cambridge University Press.