Sub-Nyquist SAR via Quadrature Compressive Sampling with Independent Measurements

Huizhang Yang [ID], **Chengzhi Chen, Shengyao Chen** * [ID] **and Feng Xi** [ID]

School of Electronic and Optical Engineering, Nanjing University of Science and Technology,
Nanjing 210094, China; hzyang@njust.edu.cn (H.Y.); chenchengzhiccz@163.com (C.C.);
xifeng@njust.edu.cn (F.X.)
*   Correspondence: chenshengyao@njust.edu.cn

check for
updates

Abstract: This paper presents an efficient sampling system for the acquisition of synthetic aperture radar (SAR) data at sub-Nyquist rate. The system adopts a quadrature compressive sampling architecture, which uses modulation, filtering, sampling and digital quadrature demodulation to produce sub-Nyquist or compressive measurements. In the sequential transmit-receive procedure of SAR, the analog echoes are modulated by random binary chipping sequences to inject randomness into the measurement projection, and the chipping sequences are independent from one observation to another. As a result, the system generates a sequence of independent structured measurement matrices, and then the resulting sensing matrix has better restricted isometry property, as proved by theoretical analysis. As a standard recovery problem in compressive sensing, image formation from the sub-Nyquist measurements has significantly improved performance, which in turn promotes low sampling/data rate. Moreover, the resulting sensing matrix has structures suitable for fast matrix-vector products, based on which we provide a first-order fast image formation algorithm. The performance of the proposed sampling system is assessed by synthetic and real data sets. Simulation results suggest that the proposed system is a valid candidate for sub-Nyquist SAR.

Keywords: synthetic aperture radar; compressive sampling; restricted isometry property; fast recovery

1. Introduction

Synthetic aperture radar (SAR) is an imaging radar mounted on a moving platform. As the platform moves, electromagnetic waves are sequentially transmitted to illuminate an area and the backscattered echoes are collected by the radar receiver to form raw data for subsequent digital processing. With increasing demands for SAR products, modern SAR systems are facing several problems. On the one hand, the raw data has resulted in a large burden on board and subsequent storage as well as downlink transmission, especially for spaceborne repeat-pass observations. Nowadays, numerous on-orbit earth observation missions are collecting SAR data continuously and have accumulated huge historical archives. On the other hand, it is often required to use large bandwidth signals for high-resolution applications, which thus need high-speed analog-to-digital converters (ADCs). These requirements present a big challenge for SAR development.

Compressive sensing (CS) theory [1], which emerged in the past decade, brings us new concepts for sub-Nyquist acquisition of SAR data. For a sparse or compressible radar scene, it is sufficient to recover a high-resolution SAR image with a small number of samples. Past investigations find that CS offers various benefits for SAR, such as reduced storage cost, increased resolvability of point targets [2], and reduced speckle [3]. In sub-Nyquist SARs, CS can be applied both in azimuth and range dimensions. Azimuth-dimensional CS is easily realized through observation at a random subset of azimuth positions [4,5], while range-dimensional CS falls into the scope of analog sub-Nyquist

sampling [6,7]. Since CS in range dimension is an important bottleneck in conventional high-resolution SARs, building analog samplers at sub-Nyquist rates is crucial for the sub-Nyquist acquisition of SAR data. In this paper, we therefore focus on the range-dimensional CS. Now, several sub-Nyquist sampling systems have been suggested for sampling radar echoes. The original sub-Nyquist sampler is random sampling [8] or non-uniform sampling [9], which digitizes analog signals at random or irregular time points. However, random/non-uniform sampling is not conducive to hardware implementation and is sensitive to jitter noise. To circumvent these disadvantages, Kirolos et al. established the random demodulation system, which modulates large bandwidth lowpass signals with a pseudo-random binary sequence, followed by lowpass filtering and low-rate ADC sampling [6]. Thereafter, Xi et al. merged the random demodulation into classical quadrature sampling and thus presented a quadrature compressive sampling (QuadCS) system to gather intermediate frequency radar echoes [10]. On the other hand, Eldar et al. provided a Xampling system [11,12] by accomplishing the random sampling in frequency domain. Among these systems, Xampling has been successfully applied to SAR systems [4]. Therefore, these sampling systems are promising in SAR data acquisition at sub-Nyquist rate.

Different from classical SARs via Nyquist sampling, sub-Nyquist SARs collect sub-Nyquist or compressive measurements, and the imaging performance greatly depends on the sensing matrix resulting from the sampling system. Since SAR observations are sequentially conducted, the sub-Nyquist measurements can be modeled as

$$\mathbf{y}_l^{cs} = \mathbf{\Phi}_l \mathbf{y}_l + \mathbf{n}_l, \quad l = 1, ..., N_a \tag{1}$$

for all N_a observations, where l is the index of azimuth position, $\mathbf{y}_l$ and $\mathbf{n}_l$ are vectors denoting the Nyquist samples of SAR echo and the sub-Nyquist measurements of observation noise, respectively, and $\mathbf{\Phi}_l$ is the measurement matrix resulting from the sampling system. Stacking all sequential measurement vectors $\mathbf{y}_l^{cs}$ into a large one, we can translate (1) into a structured CS model as

$$\mathbf{y}^{cs} = \text{diag} \left\{ \mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_{N_a} \right\} \mathbf{D} \mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{D}$ is a basis matrix for sparsely representing the Nyquist samples of SAR echo [3], $\mathbf{n}$ is the stacked noise vector, and $\mathbf{x}$ is the sparse reflecting coefficient vector. An example of the sparse reflecting coefficient is that, in a marine environment, scenes consisting of ships or islands usually lead to sparse SAR images [13,14]. In this paper, we focus on this kind of sparse scenes.

In the context of CS, the linear system (2) is typically ill-posed, and image recovery from the sub-Nyquist measurements is equivalent to a sparse recovery problem, which can be handled by many methods, such as greedy algorithms [15] and convex optimization-based algorithms [16]. The recovery errors between the ground truth and the estimates returned by these algorithms greatly depend on the sparsity of the vector $\mathbf{x}$ and the restricted isometry property (RIP) of the sensing matrix defined as $\mathbf{A} = \text{diag}\{\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_{N_a}\}\mathbf{D}$. Specifically, the matrix $\mathbf{A}$ satisfies the s-order RIP if there exists a constant $\delta_s \in (0,1)$ such that

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leqslant \|\mathbf{A}\mathbf{x}\|_2^2 \leqslant (1 + \delta_s) \|\mathbf{x}\|_2^2 \tag{3}$$

holds for all s-sparse vectors $\mathbf{x}$. In addition, the smallest δ_s is called the restricted isometry constant (RIC). If the matrix $\mathbf{A}$ satisfies $2s$-order RIP with RIC $\delta_{2s} < \delta^\#$ for some constant $\delta^\# < 1$, the recovery error between the sparse vector $\mathbf{x}$ and its estimate $\mathbf{x}^\#$ returned by those algorithms will obey

$$\left\| \mathbf{x} - \mathbf{x}^\# \right\|_2 \leq C_1 \frac{\sigma_s(\mathbf{x})_1}{\sqrt{s}} + C_2 \epsilon, \tag{4}$$

where ϵ is a upper bound of the noise norm $\|\mathbf{n}\|$. In (4), $\sigma_s(\mathbf{x})_1 = \inf_{\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_1$ denotes the best s-term approximation error of $\mathbf{x}$ in ℓ_1-norm, and C_1 and C_2 are positive constants depending on δ_{2s} (their values increase as the increase of δ_{2s}). Therefore, it is expected that the RIC δ_{2s} is

small for reasonable large s, such that the recovery error $\left\| \mathbf{x} - \mathbf{x}^{\#} \right\|_2$ is well bounded, or, in other words, the recovery is stable and robust in the sense that (4) is bounded in the presence of noise and mismodeling (for example, the scene is not reasonable sparse). Unfortunately, previous sub-Nyquist SAR works lack in RIP-based design and analysis.

The original motivation of this paper is to design a RIP-based efficient sampling system for sub-Nyqusit SAR. Here, the RIP-based efficiency means that the sensing matrix **A** has the smaller RIC, which thus provides better imaging performance with theoretical guarantees. As mentioned above, Xampling has been exploited for sub-Nyquist SAR [4], where a fixed measurement matrix is taken to sample radar echoes for all observations, i.e., the measurement matrices $\mathbf{\Phi}_l$ for $l = 1, \cdots, N_a$ are the same. However, in the context of CS, the sensing matrix with more randomness usually results in a smaller RIC, so Xampling is not in favor of an efficient sensing matrix. Along this analysis, a natural question arises: may the measurement matrix $\mathbf{\Phi}_l$ vary among different observations so that the sensing matrix **A** will have smaller RIC? In this paper, we provide an affirmative answer. By designing time-varying measurement matrices, the sampling systems that inject randomness by random modulation are competitive candidates, and they have been well studied, such as random demodulation [6]. A variant of random demodulation is QuadCS [10], as shown in Figure 1, which can effectively acquire sub-Nyquist in-phase and quadrature components of sparse bandpass signals. The QuadCS uses random modulation by chipping sequence, bandpass filtering, intermediate-frequency sampling and digital quadrature demodulation to collect sub-Nyquist measurements. It has been successfully applied to sample pulse-Doppler radar echoes with a fixed chipping sequence between different pulse repetition intervals [17–19].
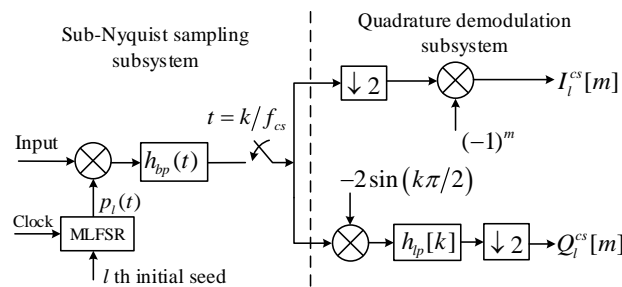

Figure 1. The architecture of QuadCS. The first subsystem is to perform low-rate random measurement; the second subsystem is a classical digital quadrature demodulation system [20] used for extract compressive in-phase and quadrature measurement.

Based on the QuadCS architecture, this paper proposes an efficient sampling system for sub-Nyquist SAR, which was partially presented in [21]. In the sequential observations, the initial seeds that control the chipping sequences are assigned to different values to generate independent chipping sequences. As a result, the measurement matrices $\mathbf{\Phi}_l$ are different from one observation to another, which is significantly different from Xampling [4]. By this strategy, more randomness is injected into the sampling system, which better meets the principle of random measurement in CS, and consequently the resulting sensing matrix has a smaller RIC. Theoretical analysis will demonstrate that the RIP attains a diversity gain N_a by this strategy, and thus the recovery performance will be improved. In addition, it is important to note that the proposed QuadCS system does not put further burden on hardware implementation because the independence of chipping sequences can be easily set beforehand by maximal-length linear feedback shift registers (MLFSR) with a set of unique initial seeds.

Another contribution of this paper is the development of a fast sparse imaging algorithm for the proposed sub-Nyquist SAR system. As an image inverse problem, the recovery of sparse images from sub-Nyquist measurements have a large problem size. Optimization methods designed for large-scale problems, such as truncated Newton interior-point method [22] and the more popular first-order methods [23–26], numerically rely on computing the matrix-vector products involving

A and its adjoint **A***. However, the large problem size precludes the direct application of these off-the-shelf optimization methods because the matrix **A** cannot be stored explicitly and it is costly, even impractical to calculate the matrix-vector products directly. In the proposed sampling system, we will decompose the matrix-vector products into several fast Fourier transform (FFT) based and Hadamard product-based fast operations. Then, the matrix **A** does not need to be stored explicitly, and the matrix-vector products have only $\mathcal{O}\left(N\log(N)\right)$ complexity and thus much less computational overhead, where N is the Nyquist sampling size of SAR echo. By exploiting the decomposition into typical first-order methods, we provide a first-order fast imaging algorithm.

The rest of this paper is organized as follows. Section 2 reviews some basic knowledge of conventional SAR and sub-Nyquist SAR. Section 3 presents our QuadCS-based sampling system for SAR. Section 4 gives a fast sparse imaging algorithm for sub-Nyquist measurements. Section 5 provides the RIP analysis of proposed system. Section 6 gives several numerical simulations to illustrate the improved performance, and finally we conclude the paper in Section 7.

Notation

The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. An integer set, $\{1, 2, \cdots, N\}$, is denoted by $[N]$. The cardinality of a set S is denoted by card(S). Vectors and matrices are denoted by lower case boldface and upper case boldface letters, e.g. **x** and **A**, respectively. The support of a vector **x** is denoted by supp(**x**). The vector output by a reconstruction algorithm is denoted by **x**$^{\#}$. For a matrix **A**, vec$(\mathbf{A})$ is the vector obtained by concatenating the columns of **A**, and diag$(\mathbf{A})$ is the diagonal matrix obtained by taking vec$(\mathbf{A})$ as its diagonal elements. The identity matrix and the zero matrix are denoted by $\mathbf{I}_N$ and $\mathbf{0}_N \in \mathbb{R}^{N \times N}$, respectively.

The real part of a complex-valued signal is denoted by Re$\{\cdot\}$. The transpose, conjugate and adjoint of a matrix are denoted by $(\cdot)^T$, $\overline{(\cdot)}$ and $(\cdot)^*$, respectively. The adjoint of an operator is also denoted by $(\cdot)^*$. The ℓ_1, Euclidean and Frobenius norm are denoted by $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$. The symbols $\otimes$ and $\circ$ denote the Kronecker product and the Hadamard product, respectively.

2. SAR Imaging and Sub-Nyquist SAR Model

In this section, we give a basic introduction to SAR imaging, and then discuss the sub-Nyquist SAR and associated image formation method.

2.1. SAR Imaging

For simplicity, we take the stripmap SAR as an example. As shown in Figure 2, the radar platform travels along azimuth direction at a constant velocity while the radar illuminates a scene. As the radar moves, the SAR transmits a pulse Re $\left\{ s\left(t\right)e^{j2\pi f_0 t} \right\}$ with carrier frequency f_0 and bandwidth B to illuminate the scene and collect its echo data, and this procedure is sequentially conducted every pulse repetition interval (PRI) T to collect multiple coherent measurements that contains the reflectivity information about the scene. Supposing that the illuminated scene is discretized into K scattering units with predesigned resolution, then the received backscattered signal can be modeled as superpositions of the echo signals due to these scattering units. For the l-th observation, after demodulation, the complex baseband envelope can be expressed as

$$y_l\left(t\right) = \sum_{n=1}^{K} D_{l,n}\left(t\right)\sigma_n. \tag{5}$$

In (5), σ_n is the complex reflectivity coefficient of the n-th scattering unit and $D_{l,n}\left(t\right)$ is a waveform related to the scattering unit as

$$
\begin{aligned}
D_{l,n}\left(t\right) =& w_a\left(lT - \eta_n\right)s\left(t - 2R_{l,n}/c\right) \\
& \times \exp\left(-j4\pi f_0 R_{l,n}/c\right),
\end{aligned} \tag{6}
$$

where $w_a(\cdot)$ is the antenna beam pattern, η_n is the beam center crossing time, and $R_{l,n}$ is the slant range of the n-th scattering unit. For subsequent digital processing, the baseband echo (6) is sampled according to the Nyquist theorem to collect its complex sample vector $\mathbf{y}_l$ with length $N_r = BT$. A coherent combination of N_a echo data forms a two-dimensional data matrix of complex samples $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_{N_a}]$. Let us denote the reflectivity map of the discretized scene by a complex image matrix $\mathbf{X}$ with size $K_r \times K_a$. By stacking the echo data matrix and image matrix into vectors respectively, we obtain the following observation model

$$\mathbf{y} = \mathbf{D}\mathbf{x}, \tag{7}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$ is the vectorized Nyquist rate echo data with length $N = N_r N_a$, and $\mathbf{x} = \text{vec}(\mathbf{X})$ is the image vector with length $K = K_r K_a$. The SAR image can be well focused by traditional matched-filter-based focusing methods, $\mathbf{x}_{mf} = \mathbf{D}^* \mathbf{y}$, such as range-Doppler algorithm and chirp scaling algorithm (CSA). A virtue of these algorithms is that they are numerically efficiency with $\mathcal{O}(N \log(N))$ complexity, owing to range and azimuth decoupled processing and fast Fourier transform (FFT) operations.
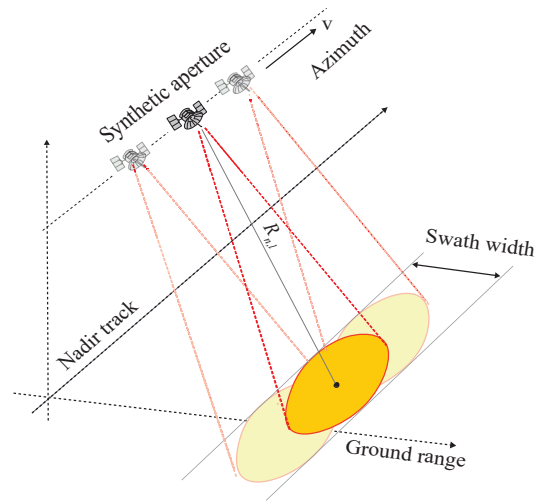

Figure 2. Illustration of stripmap SAR imaging geometry.

2.2. Sub-Nyquist SAR

Since its invention, CS theory has been applied to SAR systems to reduce the data rate and subsequent demands on storage [27,28]. In classical SAR systems, radar echoes are sampled at Nyqusit rate in azimuth and range dimensions, respectively. Therefore, CS can be conducted both in azimuth and range dimensions. In the azimuth dimension, CS is often realized through randomly transmitting a reduced number of pulses at irregular azimuth positions. A shortcoming of this strategy is that the underlying mathematical model falls into the paradigm of block sampling, which limits the performance of its sensing matrix. In the range dimension, CS of the analog radar signals involves structured sampling systems, such as the Xampling system reported in [4,12]. This paper focuses on the sub-Nyquist sampling in range dimension, so the azimuth-dimension CS is not considered. Our work can be naturally extended to the case of azimuth-dimension CS.

In a sub-Nyquist SAR, the sub-Nyquist measurements of the l-th observation can be expressed as

$$\mathbf{y}_l^{cs} = \mathbf{\Phi}_l \mathbf{y}_l + \mathbf{n}_l, \quad l \in [N_a], \tag{8}$$

where $\mathbf{y}_l^{cs}$ and $\mathbf{n}_l$ are the length-M_r sub-Nyquist measurement vector and the measurement noise vector, respectively, and $\mathbf{\Phi}_l$ is the $M_r \times N_r$ measurement matrix specified by the sampling system for l-th observation. Let us collect the sample vectors to form a 2D measurement $\mathbf{Y}^{cs} = \left[\mathbf{y}_1^{cs}, \cdots, \mathbf{y}_{N_a}^{cs}\right]$,

and denote its vectorization by $\mathbf{y}^{cs} = \text{vec}\left(\mathbf{Y}^{cs}\right)$ with length $M = M_r N_a$. We can express the sub-Nyquist measurements of (7) as

$$\mathbf{y}^{cs} = \mathbf{Ax} + \mathbf{n}, \tag{9}$$

where $\mathbf{A} = \text{diag}\left(\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{N_a}\right)\mathbf{D} \in \mathbb{C}^{M \times K}$ is the sensing matrix and $\mathbf{n} = \text{vec}([\mathbf{n}_1, \cdots, \mathbf{n}_{N_a}]) \in \mathbb{C}^{M \times 1}$ is the contaminating measurement noise.

Typically, the measurement dimension M is much smaller than the image size K, and then image formation from the linear system (9) is an ill-posed problem. However, assuming that the image vector $\mathbf{x}$ is sparse and the sensing matrix $\mathbf{A}$ satisfies the RIP, the image can be estimated by solving the following basis pursuit denoising (BPDN) problem

$$\min_{\mathbf{x}} \frac{1}{2} \left\|\mathbf{Ax} - \mathbf{y}^{cs}\right\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{10}$$

where $\lambda > 0$ is a regularization parameter used to trade off between sparsity and data fidelity. For this optimization problem, the very high dimension of $\mathbf{x}$ precludes the direct application of classical optimization methods, such as inner-point methods and greedy algorithms. However, matrix-vector products involving $\mathbf{A}$ and its adjoint $\mathbf{A}^*$, if they exist, can be used to speed up many popular algorithms for solving (10), such as *fast iterative shrinkage-thresholding algorithm* (FISTA) [25]. In this sense, the sampling system is expected to have special structures for fast matrix-vector products. Then, the recovery algorithm can be adapted for fast image formation.

3. Quadrature Compressive Sampling for SAR with Independent Measurements

In this section, we present the QuadCS-based sampling system for sub-Nyquist SAR. For sequential SAR observations, we use independent chipping sequences to modulate the echoes; thus, the measurement matrix $\boldsymbol{\Phi}_l$ is different from one to another. We call this strategy as independent measurements. Then, we provide the frequency-domain representation of this strategy for the convenience of subsequent RIP analysis. As will be shown in Section 5, this strategy will lead to smaller RICs and therefore better recovery performance.

3.1. Sampling SAR Echoes via QuadCS

The fundamentals of the QuadCS system [10] are shown in Figure 1, which consists of two subsystems: a sub-Nyquist sampling subsystem and a digital quadrature demodulation subsystem. Before inputting them into the system, the radar signal is down converted to intermediate frequency (IF) f_I. With an IF input signal, the first subsystem generates a compressive bandpass signal and its bandpass samples, and then the second subsystem translates the bandpass samples into the complex baseband samples by quadrature demodulation.

Let us take the l-th observation to illustrate its operating principle. In the first subsystem, the IF radar echo $\text{Re}\left\{y_l(t)e^{j2\pi f_I t}\right\}$ is modulated by a chipping sequence $p_l(t)$ with the bandwidth not less than that of $y_l(t)$. The mixing operation will spread the frequency content of the baseband signal $y_l(t)$ to full spectrum of $p_l(t)$. The mixing output is filtered by a bandpass filter $h_{bp}(t)$ with center frequency f_0 and bandwidth $B_{cs} \ll B$. The filter outputs the compressive bandpass signal $r_l(t)$,

$$\begin{aligned} r_l(t) &= \int_{-\infty}^{\infty} h_{bp}\left(t - \tau\right) p_l\left(\tau\right)\text{Re}\left\{y_l(\tau)e^{j2\pi f_I \tau}\right\}d\tau \\ &= \text{Re}\left\{y_l^{cs}\left(t\right)e^{j2\pi f_I t}\right\}, \end{aligned} \tag{11}$$

where $y_l^{cs}(t)$ is the compressive complex envelope given by

$$y_l^{cs}(t) = \int_{-\infty}^{\infty} h_{bp}\left(t - \tau\right) e^{-j2\pi f_I(t-\tau)} p_l\left(\tau\right) y_l\left(\tau\right)d\tau, \tag{12}$$

with $I_l^{cs}(t) = \mathrm{Re}\left\{y_l^{cs}(t)\right\}$ and $Q_l^{cs}(t) = \mathrm{Im}\left\{y_l^{cs}(t)\right\}$ denoting the *compressive I* and *Q* components of $y_l^{cs}(t)$, respectively. The filter output $y_l^{cs}(t)$ is then sampled by a low-rate ADC to generate a sub-Nyquist sequence $y_l^{cs}[k] = r_l(k/f_{cs})$, which is the baseband sample of $y_l^{cs}(t)$. The sampling rate f_{cs} is set according to bandpass sampling theorem as $f_{cs} = (4f_L + 2B_{cs})/(4k+1)$, where $f_L = f_I - B_{cs}/2$ and k is a positive integer satisfying $k \leq \lfloor f_L/2B_{cs}\rfloor$. The second subsystem is to extract digital *I* and *Q* sequences from the sub-Nyquist sequence $y_l^{cs}[k]$. Its operation is the same as classic quadrature sampling [20]. Because of the down-sampling operation, the rate of the digital *I* and *Q* sequences, $I_l^{cs}[m] = I_l^{cs}(mT_{cs})$ and $Q_l^{cs}[m] = Q_l^{cs}(mT_{cs})$, is half that of $y_l^{cs}[k]$, $T_{cs} = 2/f_{cs}$. In a PRI, we obtain $M_r = \lfloor T/T_{cs}\rfloor$ sub-Nyquist complex samples $y_l^{cs}[m] = I_l^{cs}[m] + jQ_l^{cs}[m]$ or $2M_r$ samples of *I* and *Q* components, which is much less than $2BT$ by the classical Nyquist sampling.

Although the QuadCS system works on analog bandpass signals, its output $y_l^{cs}[m]$ can be characterized as a linear combination of the sparse coefficient vector $\mathbf{x} = [\sigma_1, \cdots, \sigma_N]^T$. Inserting (5) and (6) into (12), we have

$$
\begin{aligned}
y_l^{cs}(t) &= \sum_{n=1}^{K} \sigma_n \int_{-\infty}^{\infty} h_{bp}(t-\tau) e^{-j2\pi f_I(t-\tau)} \\
&\quad \times p_l(\tau) D_{l,n}(\tau)\, d\tau \\
&= \sum_{n=1}^{K} \sigma_n A_{l,n}(t),
\end{aligned}
\tag{13}
$$

where $A_{l,n}(t)$ is the compressive signal of $D_{l,n}(t)$

$$
A_{l,n}(t) = \int_{-\infty}^{\infty} h_{bp}(t-\tau) e^{-j2\pi f_I(t-\tau)} p_l(\tau) D_{l,n}(\tau)\, d\tau.
\tag{14}
$$

Then, the sub-Nyquist complex samples can be expressed as

$$
y_l^{cs}[m] = \sum_{n=1}^{K} \sigma_n A_{l,n}(mT_{cs}).
\tag{15}
$$

To get a standard CS formulation, let us define the sub-Nyquist measurements vector of the *l*-th observation as

$$
\mathbf{y}_l^{cs} = \left[y_l^{cs}[0], \cdots, y_l^{cs}[M_r-1]\right]^T,
\tag{16}
$$

collect all N_a observations as $\mathbf{Y}^{cs} = \left[\mathbf{y}_1^{cs}, \mathbf{y}_2^{cs}, \cdots, \mathbf{y}_{N_a}^{cs}\right]$ or its vector form $\mathbf{y}^{cs} = \mathrm{vec}\left(\mathbf{Y}^{cs}\right)$, and define the sensing matrix $\mathbf{A}$ as

$$
(\mathbf{A})_{m+(l-1)M_r,n} = A_{l,n}(mT_{cs}),
\tag{17}
$$

for $m \in [M_r]$, $l \in [N_a]$ and $n \in [N]$. Then, we have the sub-Nyquist measurements model $\mathbf{y}^{cs} = \mathbf{A}\mathbf{x}$ as given by (9) in the noisy case. However, the structure of the sensing matrix $\mathbf{A}$ here is not as clear as that in (2). To decompose the sensing matrix into an explicit structured form, we next analyze the QuadCS in frequency domain, which is also important for subsequent discussions on the fast imaging algorithm and RIP.

3.2. Frequency-Domain Representation

From (11), we can see that the QuadCS involves a time-domain product and a bandpass filtering operation. The time-domain product is equivalent to frequency-domain convolution, which can be described by the product of a Toeplitz matrix by a vector, and the bandpass filtering can be modelled by a truncation operation in frequency domain under the assumption of idea frequency response. Based on the above-mentioned view, we have the frequency-domain measurement model of QuadCS as studied in [10].

Since each observation interval has a duration T, the chipping sequences $p_l(t)$ can be expressed in Fourier series as

$$p_l(t) = \sum_{i=-L_p}^{L_p} \rho_{i,l} e^{j2\pi f_p it},$$ (18)

where $f_p = 1/T$, $\rho_{i,l}$ is the Fourier coefficient of $p_l(t)$, and $L_p \geq BT$ is a positive integer. Denoting the discrete Fourier transform (DFT) of $\mathbf{y}_l^{cs}$ by $\hat{\mathbf{y}}_l^{cs}$, it can be given by [29]

$$\hat{\mathbf{y}}_l^{cs} = \sqrt{\frac{1}{M_r}} \mathbf{T}_l \mathbf{F}_r \mathbf{y}_l,$$ (19)

where $\mathbf{F}_r$ is a $N_r \times N_r$ normalized DFT matrix after fftshift (fftshift is to shift zero-frequency component to the center of spectrum), and $\mathbf{T}_l$ is a $M_r \times N_r$ Toeplitz matrix defined by the Fourier coefficients as

$$\mathbf{T}_l = \begin{bmatrix} \rho_{-L_0+N_r-1,l} & \rho_{-L_0+N_r-2,l} & \cdots & \rho_{-L_0,l} \\ \rho_{-L_0+N_r,l} & \rho_{-L_0+N_r-1,l} & \cdots & \rho_{-L_0+1,l} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{L_0,l} & \rho_{L_0-1,l} & \cdots & \rho_{-L_0+M_r-1,l} \end{bmatrix},$$ (20)

where $L_0 = M_r + N_r - 1$. The Toeplitz structure of $\mathbf{T}_l$ results from the convolution between the spectrums of the chipping sequence and the input signal. Due to bandpass filtering, $\mathbf{T}_l$ does not contain all Fourier coefficients of $p_l(t)$. Next, according to (19), we can describe the time-domain sub-Nyquist measurements as

$$\mathbf{y}_l^{cs} = \mathbf{F}_m^* \hat{\mathbf{y}}_l^{cs} = \sqrt{\frac{1}{M_r}} \mathbf{F}_m^* \mathbf{T}_l \mathbf{F}_r \mathbf{y}_l,$$ (21)

where $\mathbf{F}_m$ is a $M_r \times M_r$ normalized DFT matrix after fftshift. Then, the measurement matrix $\boldsymbol{\Phi}_l$ can be described as

$$\boldsymbol{\Phi}_l = \sqrt{\frac{1}{M_r}} \mathbf{F}_m^* \mathbf{T}_l \mathbf{F}_r,$$ (22)

which is independent from one observation to another due to the adoption of different chipping sequences. Finally, according to (8) and (9), we can rewrite the sensing matrix as a structured form, $\mathbf{A} = \text{diag}(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_{N_a}) \mathbf{D}$. The first factor of $\mathbf{A}$ is a block diagonal matrix with independent diagonal blocks. Therefore, the sensing matrix $\mathbf{A}$ has more randomized structures, which better meets the principle of random measurement in CS. In addition, the block diagonal and Toeplitz structures are also helpful for deriving a fast sparse recovery algorithm, as will be shown later.

3.3. Remarks

The QuadCS generates independent measurement matrices while does not complicate hardware implementation. In the system, chipping sequences are realized by a pseudo-random number generator, which can be implemented using a MLFSR. The MLFSR has the benefit of providing a random sequence of ±1 with zero average, and it offers the ability of regenerating the same sequence by a given initial seed [6]. This technique has been widely used in modern radio communication, such as CDMA. During sequential observations of SAR, the initial seeds are easily assigned to different values, which can be set in advance or in real-time. On the contrary, in the Xampling system [4,12], it is not allowed to generate varied measurement matrices among different observations, because four bandpass filters with different passbands are utilized to perform sub-Nyquist measurement and the measurement matrix is inherently fixed when the hardware is built. Moreover, bandpass filters cannot sample SAR echoes randomly enough in frequency domain and thus often generate bad sensing matrices, as observed in the following simulations. As for non-uniform sampling, it often imposes strict timing requirements on the digital circuits to preserve timing information [8,30], which makes it

difficult to interface with synchronous digital circuits. In fact, it is easier to build a high-rate modulator via mixer/chipping sequence combination than a specially designed non-uniform sampler [31].

4. Fast Sparse Imaging with QuadCS Measurements

As a standard sparse recovery problem in CS, we model our imaging problem as the BPDN formulation in (10). For the large-scale problem, the very high dimensions of $\mathbf{A}$ often precludes the direct application of off-the-shelf optimization methods because of two limitations: the matrix $\mathbf{A}$ cannot be stored explicitly, and the matrix-vector products involving $\mathbf{A}$ and its adjoint $\mathbf{A}^*$ cannot be quickly calculated. If these two limitations are well resolved, several popular solvers can be used to solve (10), such as FISTA [32], SpaRSA [33], TFOCS [26] and L1LS [22]. In the proposed sampling system, we exploit the block and Toeplitz structure of the sensing matrix $\mathbf{A}$ and decompose the matrix-vector products into several fast operations. Then, the matrix $\mathbf{A}$ does not need to be stored explicitly, and the matrix-vector products can be fast calculated. By inserting the decomposition into above-mentioned solvers, we derive a FISTA-based fast SAR imaging algorithm for the proposed sub-Nyquist SAR system, and analyze the algorithm complexity.

4.1. Fast Matrix-Vector Products

As we have discussed, the sensing matrix $\mathbf{A} = \mathbf{\Phi}D$ consists of two parts: measurement matrix $\mathbf{\Phi}$ and basis matrix $\mathbf{D}$. To obtain the desired fast matrix-vector products for $\mathbf{A}$ and $\mathbf{A}^*$, we are going to exploit the inner structures of the two matrices. The following content is to expand the derivations in detail.

Let us first discuss the Toeplitz structure as shown in the frequency-domain model (21). Since $\mathbf{T}_l$ is a Toeplitz matrix, it can be viewed as a submatrix of a $(2L_0 + 1) \times (2L_0 + 1)$ circulant matrix $\mathbf{C}_l$ defined as

$$\mathbf{C}_l = \begin{bmatrix} \rho_{L_0, l} & \rho_{L_0-1, l} & \cdots & \rho_{-L_0, l} \\ \rho_{-L_0, l} & \rho_{L_0, l} & \cdots & \rho_{-L_0+1, l} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{L_0-1, l} & \cdots & \rho_{-L_0, l} & \rho_{L_0, l} \end{bmatrix}. \tag{23}$$

Specifically, $\mathbf{T}_l$ is exactly the right N_r columns and top M_r rows of $\mathbf{C}_l$. Let us define row and column truncation matrices as

$$\mathbf{\Delta}_r = \begin{bmatrix} \mathbf{I}_{M_r} & \mathbf{0}_{N_r-1} \end{bmatrix}, \tag{24}$$

and

$$\mathbf{\Delta}_a = \begin{bmatrix} \mathbf{0}_{M_r-1} \\ \mathbf{I}_{N_r} \end{bmatrix}, \tag{25}$$

respectively. Then, we embed $\mathbf{T}_l$ into $\mathbf{C}_l$ as follows:

$$\mathbf{T}_l = \mathbf{\Delta}_r \mathbf{C}_l \mathbf{\Delta}_a. \tag{26}$$

Since circulant matrices can be diagonalized by DFT matrices, there exist fast matrix-vector products for circulant matrices via FFT and the computational complexity is $\mathcal{O}(N\log(N))$. The goal of (26) is to extend this property to the Toeplitz matrix $\mathbf{T}_l$. Specifically, let us define an integer $L = 2L_0 + 1$, and a vector $\boldsymbol{\rho}_l = \begin{bmatrix} \rho_{L_0, l}, \rho_{-L_0, l}, \cdots, \rho_{L_0-1, l} \end{bmatrix}^T$ which is the first column vector of $\mathbf{C}_l$, and denote the $L \times L$ DFT matrix as $\mathbf{F}_L$. Then, the circulant matrix $\mathbf{C}_l$ can be diagonalized as

$$\mathbf{C}_l = \mathbf{F}_L \text{diag}\left(L\boldsymbol{\theta}_l\right) \mathbf{F}_L^*, \tag{27}$$

where $\boldsymbol{\theta}_l = \mathbf{F}_L \boldsymbol{\rho}_l$. Inserting (26) and (27) into (21), we have

$$\mathbf{y}_l^{cs} = \sqrt{\frac{L}{M_r}} \mathbf{F}_m^* \mathbf{\Delta}_r \mathbf{F}_L \mathrm{diag}\left(\boldsymbol{\theta}_l\right) \mathbf{F}_L^* \mathbf{\Delta}_a \mathbf{F}_r \mathbf{y}_l. \tag{28}$$

Defining $\mathbf{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{N_a}]$, then the 2D QuadCS measurement $\mathbf{Y}^{cs}$ can be written as

$$\mathbf{Y}^{cs} = \sqrt{\frac{L}{M_r}} \mathbf{F}_m^* \mathbf{\Delta}_r \mathbf{F}_L \left(\mathbf{\Theta} \circ \mathbf{F}_L^* \mathbf{\Delta}_a \mathbf{F}_r \mathbf{Y}\right), \tag{29}$$

which mainly involves FFTs and thus has a low complexity. Note that, from (29), we find that the random matrix $\mathbf{\Theta}$ performs "Hadamard scrambling" to inject randomness into the measurements.

Next, we analyze the inner structure of the basis matrix $\mathbf{D}$. It has been suggested that conventional imaging algorithms, such as CSA, can be exploited to derive its fast matrix-vector product for $\mathbf{D}$ [5,34], which demonstrate that the matrix is also highly structured. We assume that the SAR operates in stripmap mode and the transmitted pulses are linear frequency modulated (LFM); then, CSA can be used to obtain the SAR image, which can be described in a compact form as

$$\mathbf{X}_{CSA} = \left(\mathbf{H}_3 \circ \mathbf{F}_r^* \left(\mathbf{H}_2 \circ \mathbf{F}_r \left(\mathbf{H}_1 \circ \mathbf{Y} \mathbf{F}_a\right)\right)\right) \mathbf{F}_a^*. \tag{30}$$

In this equation, $\mathbf{X}_{CSA}$ is the SAR image output by CSA, $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$ are phase compensation matrices (i.e., their entries lie in the unit circle in $\mathbb{C}$, for more details, cf. chapter 6 in [35]), and $\mathbf{F}_r$, $\mathbf{F}_a$ are the range and azimuth DFT matrices, respectively. Because CSA is a high precision algorithm, $\mathbf{X}_{CSA}$ is a valid approximation of the reflectivity map $\mathbf{X}$. For ease of discussion, we omit the approximation error, i.e., we take $\mathbf{X} = \mathbf{X}_{CSA}$, and in this sense, $\mathbf{X}$ and $\mathbf{Y}$ have the same size, i.e., $K_r = N_r$ and $K_a = N_a$. After simple derivation, the inversion of (30) is given by [5]

$$\mathbf{Y} = \left(\overline{\mathbf{H}}_1 \circ \mathbf{F}_r^* \left(\overline{\mathbf{H}}_2 \circ \mathbf{F}_r \left(\overline{\mathbf{H}}_3 \circ \mathbf{X} \mathbf{F}_a\right)\right)\right) \mathbf{F}_a^*. \tag{31}$$

It is easy to see that the linear system (31) is equivalent to $\mathbf{y} = \mathbf{D}\mathbf{x}$ with proper reshape operations. Furthermore, since the involved FFTs and Hadamard products in (30) and (31) are unitary operators, the matrix $\mathbf{D}$ is also unitary [5].

4.2. FISTA-Based Fast Imaging Algorithm

FISTA is an accelerated variant of the iterative shrinkage-thresholding algorithms (ISTA), and it has a quadratic rate of convergence thanks to the generalized Nesterov's method [23,25]. Each iteration of FISTA involves matrix-vector product involving $\mathbf{A}$ and its adjoint $\mathbf{A}^*$ followed by a shrinkage/soft-threshold step as follows:

$$\mathbf{x}_{k+1} = \mathcal{T}_{\lambda/L_f}\left(\mathbf{z}_k - \mathbf{A}^* \left(\mathbf{A}\mathbf{z}_k - \mathbf{y}_{cs}\right)/L_f\right), \tag{32}$$

where $\mathbf{x}_{k+1}$, $\mathbf{z}_k$ are iterative variables, L_f is the Lipschitz constant for the gradient Δf of the function $f\left(\mathbf{x}\right) = \|\mathbf{A}\mathbf{x} - \mathbf{y}^{cs}\|_2^2$, and $\mathcal{T}_{\lambda/L_f}\left(\cdot\right)$ is the soft shrinkage operator defined as

$$\mathcal{T}_\tau(\mathbf{x})_i = \begin{cases} (|\mathbf{x}_i| - \tau) \frac{\mathbf{x}}{|\mathbf{x}_i|}, & |\mathbf{x}_i| > \tau; \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

From (32), we know that the matrix-vector products involving $\mathbf{A}$ and its adjoint $\mathbf{A}^*$ dominate the computational cost in each iteration. The previous subsection shows that the matrix-vector products can be fast calculated through matrix decomposition. Now, we merge this calculation strategy into FISTA and thus provide a FISTA-based fast imaging algorithm, as shown in Algorithm 1.

Algorithm 1 Fast sparse imaging with QuadCS measurements

Input:

    Operators $\mathcal{M}$, $\mathcal{D}$ and their adjoints; QuadCS measurement $\mathbf{Y}^{cs}$; number of iterations *Niter*; λ;

Output:

    Sparse SAR image $\mathbf{X}^{\#}$;

1: initial: take λ, $L_f \in \mathbb{R}$, $\mathbf{Z}_0 = \mathbf{X}_0 = \mathbf{0}$, $q_0 = 1$, $k = 1$;

2: **repeat**

3:    $\mathbf{Z}_k = \mathbf{Z}_{k-1} - \frac{1}{L_f} \mathcal{D}^* \mathcal{M}^* \left(\mathcal{M} \mathcal{D} \mathbf{Z}_{k-1} - \mathbf{Y}^{cs} \right)$;

4:    $\mathbf{X}_k = \mathcal{T}_{\lambda/L_f} \left(\mathbf{Z}_k \right)$;

5:    $q_k = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 4q_{k-1}^2}$;

6:    $\mathbf{Z}_k = \mathbf{X}_k + \frac{q_{k-1}-1}{q_k} (\mathbf{X}_k - \mathbf{X}_{k-1})$;

7:    $k = k + 1$;

8: **until** $k \geq$ *Niter*;

9: return $\mathbf{X}^{\#} = \mathbf{X}_{k-1}$;

For ease of description, we denote the right sides of (29) and (31) by operators as $\mathcal{M}\mathbf{Y}$ and $\mathcal{D}\mathbf{X}$, respectively, and denote vec($\cdot$) by an operator $\mathcal{V}$. With these notations, we can express the matrix-vector product of $\mathbf{A}$ as

$$\mathbf{A}\mathbf{x} = \mathcal{V}\mathcal{M}\mathcal{D}\mathcal{V}^*\mathbf{x}. \tag{34}$$

Similarly, the matrix-vector product of the adjoint $\mathbf{A}^*$ is given by

$$\mathbf{A}^*\mathbf{y}^{cs} = \mathcal{V}\mathcal{M}^*\mathcal{D}^*\mathcal{V}^*\mathbf{y}^{cs}, \tag{35}$$

where the adjoint operator $\mathcal{D}^*$ is given by the operations on the right side of (30) and $\mathcal{M}^*$ is given by

$$\mathcal{M}^*\mathbf{Y}^{cs} = \sqrt{\frac{L}{M_r}} \, \mathbf{F}_r^* \boldsymbol{\Delta}_a^T \mathbf{F}_L \left(\overline{\boldsymbol{\Theta}} \circ \mathbf{F}_L^* \boldsymbol{\Delta}_r^T \mathbf{F}_m \mathbf{Y}^{cs} \right). \tag{36}$$

It is clear that the involved operations in the right sides of (34) and (35) mainly consist of FFTs and Hadamard products. Therefore, we can conclude that $\mathbf{A}$ and its adjoint $\mathbf{A}^*$ both admit fast matrix-vector products.

4.3. Complexity

Without fast matrix-vector products for $\mathbf{A}$ and $\mathbf{A}^*$, we need to store the $M \times N$ matrix $\mathbf{A}$ explicitly and compute the involved matrix-vector products directly. For a small scene of 1024×1024 pixels and compression ratio $M/N = 1/8$, it requires 1024 GB to store the matrix $\mathbf{A}$ with double precision, and roughly 1.3744×10^{11} complex multiplications to compute the matrix-vector product $\mathbf{A}\mathbf{x}$. Obviously, the image recovery will be numerically intractable. With the fast matrix-vector multiplications (34) and (35), we only need to store $N_r \times N_a$ matrices $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$ and $\boldsymbol{\Theta}$. It only requires 32 MB to store the four matrices and roughly 8.2×10^4 complex multiplications to compute $\mathbf{A}\mathbf{x}$ for a scene of 1024×1024 size, which are significantly smaller than the previous requirements. In general, the storage cost is reduced form $\mathcal{O}(MN)$ to $\mathcal{O}(N)$, and the computational cost is reduced from $\mathcal{O}(MN)$ to $\mathcal{O}(N\log(N))$ per iteration, respectively.

Looking back to the traditional imaging algorithm, CSA, it only needs to evaluate (30) once to form a SAR image. Specifically, it involves $2N_r N_a \left(\log\left(N_r \right) + \log\left(N_a \right) \right) + 3N_r N_a$ complex products, or in a compact form, $\mathcal{O}\left(N \log\left(N \right) \right)$ complexity. The proposed iterative algorithm exhausts more overhead compared to CSA, depending on the number of iterations. Indeed, the essence of CS is to sacrifice the computational cost in return for the reduction of sampling/data rate.

5. Restricted Isometry Property (RIP) Analysis

In this section, we derive the RIP of the sensing matrix **A** resulting from the proposed sampling system. The established RIP condition shows that the strategy of independent measurements provides a "diversity gain", which is up to N_a. Therefore, the sensing matrix **A** has smaller RICs with independent measurements compared to that with fixed measurements (i.e., with equal chipping sequences), and thus leads to lower sampling/data rate and smaller recovery error.

Our derivation is based on the concentration of measure (CoM) inequality [36], which is one of the leading techniques for the RIP analysis of random sensing matrices in CS. The CoM inequalities have been established for the random matrix with i.i.d. entries and other structured matrices. However, these results are not applicable to the proposed sensing matrix **A** because of its special factornizatio. In the following, we first derive a CoM inequality of the matrix **A**, and then establish a RIP condition on it. The CoM inequality of **A** is formulated in Lemma 1.

Lemma 1. *Let* **x** *be a s-sparse vector in* $\mathbb{C}^N$. *For an arbitrary* $0 < \delta < 1$, *we have*

$$\mathbb{P}\left(\left|\|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2\right| \ge \delta \|\mathbf{x}\|_2^2\right) \le 4\,exp\left(-\frac{\delta\xi\varpi^2 M}{16\chi s}\right), \tag{38}$$

where M is the number of sub-Nyquist measurements, χ *denotes two settings of chipping sequences and is defined as*

$$\chi = \begin{cases} 1, & \text{independent chipping sequences,} \\ N_a, & \text{equal chipping sequences.} \end{cases} \tag{39}$$

ξ *is the duty cycle of radar transmitted signals, and* ϖ *is the maximum value of the normalized antenna beam pattern defined as*

$$\varpi = \frac{w_a\,(0)\,T_a}{\sqrt{\int_{-T_a/2}^{T_a/2} w_a^2\,(\eta)\,d\eta}}, \tag{40}$$

where T_a *is the antenna footprint time.*

Proof. The proof is given in Appendix A.1. □

Lemma 1 points out that, with independent measurements, the probabilistic tail bounds of the CoM decays exponentially N_a times faster than that with fixed measurements. Since the CoM inequalities quantify how well a random matrix preserve the norm of a high-dimensional signal in the mapped low-dimensional space [37], the independent measurements result in a better distance-preserving property for the sensing matrix **A**, which is the essence of RIP.

Based on the CoM inequality in Lemma 1, we can derive a uniform tail bound of $\left\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}_s\right\|_2$ for a fixed subset $S \subset [N]$ with card $(S) = s$ using the covering number estimate [38]. Then, we take the union bound over all subsets $S \subset [N]$ of cardinality s, to derive the following RIP condition.

Theorem 1. *The restricted isometry constant* δ_s *of the sensing matrix* **A** *satisfies* $\delta_s < \delta$ *for some constant* $0 < \delta < 1$ *with probability exceeding* $1 - \eta$ *provided that*

$$M \ge \frac{32\chi s}{\delta\xi\varpi^2}\left(s\ln\left(\frac{eN}{s}\right) + 4s\ln(3) + ln\left(\frac{\eta}{4}\right)\right). \tag{41}$$

Proof. The proof is given in Appendix A.2. □

Theorem 1 gives the RIP condition on the sensing matrix **A**. The key parameters in (41) are M and χ, which describe the measurement amount and the setting of chipping sequences, respectively. If M (or equivalently, compression ratio) is properly chosen according to the sparsity s and the imaged scene size N, the sensing matrix **A** will satisfy RIP with high probability in both settings of

chipping sequences. However, the measurement amounts required by the two settings vary somewhat surprisingly, up to a multiplicative factor N_a. This variance suggests that the strategy of independent measurements provides a diversity gain, which is similar to the waveform diversity gain in MIMO radar. For a given imaged scene, the diversity gain results in lower sampling/data rates. On the other hand, the parameters ξ and ω are often fixed in a given SAR operation mode, and they suggest that the duty cycle and the antenna beam pattern have a direct influence on the RIP. Intuitively, for smaller ξ or ω, each column of the basis **D** will be localized in smaller regions, which is bad for the sensing matrix.

Another benefit of the diversity gain is that, with the same number of measurements, the RIC is expected to be much smaller, as stated in the following corollary.

Corollary 1. *Let δ_s and γ_s be the s-order RICs of the sensing matrices for the setting of equal and independent chipping sequences, respectively. For fixed M and η, the RICs satisfy $\delta_s < \delta$ and $\gamma_s < \delta/N_a$ simultaneously for some constant $0 < \delta < 1$ with probability exceeding $1 - \eta$.*

Replacing δ_s by δ_s/χ in Theorem 1 will prove the corollary. According to CS theory, smaller RIC means smaller recovery error. Thus, independent measurements have better performance than fixed measurements, as revealed by Theorem 2.

Theorem 2. *For a fixed number of measurements M and a fixed probabilistic guarantee $1 - \eta$, let s_1 and s_2 be the maximum integers that satisfy $\delta_{2s_1} < \delta^\#$ and $\gamma_{2s_2} < \delta^\#$ with equal and independent chipping sequences, respectively, where $\delta^\# = \sqrt{2} - 1$. Let $\mathbf{x}_1^\#$ and $\mathbf{x}_2^\#$ be the solutions of the constrained BPDN problem*

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}^{cs}\| \leq \epsilon \tag{42}$$

in the two settings of chipping sequences (39), respectively, then

$$
\begin{aligned}
\left\|\mathbf{x} - \mathbf{x}_2^\#\right\|_2 &\leq C_1 \frac{\sigma_{s_2}(\mathbf{x})_1}{\sqrt{s_2}} + C_2\epsilon \\
&\approx C_1 \frac{\sigma_{\sqrt{N_a}s_1}(\mathbf{x})_1}{\sqrt{\sqrt{N_a}s_1}} + C_2\epsilon \\
&< C_1 \frac{\sigma_{s_1}(\mathbf{x})}{\sqrt{s_1}} + C_2\epsilon ,
\end{aligned}
\tag{43}
$$

where the last line is exactly the error bound of $\|\mathbf{x} - \mathbf{x}_1^\#\|_2$.

Proof. The proof is shown in Appendix A.3. □

Theorem 2 makes a comparison between the recovery error bounds about the two settings of chipping sequences. In (43), the error bound consists of two parts: one is the best *s*-term approximation error and the other comes from the measurement noise. As shown in Appendix A.3, we have a gain factor $s_2/s_1 \approx \sqrt{N_a}$, which results in much lower best *s*-term approximation error as shown in (43). Therefore, we can recover more sparse targets in the case of high signal-to-noise ratio (SNR). For example, when $N_a = 4096$, the gain factor will be 64. Thus, the stability of image recovery is significantly improved provided that the scene is not reasonable sparse. When the SNR decreases, the noisy contribution $C_2\epsilon$ gradually dominates the error bounds, and consequently, the relative errors get closer. Namely, the advantage of QuadCS with independent measurement is weakened in this case. However, there is no doubt that $\mathbf{x}_2^\#$ always has a smaller recovery error than $\mathbf{x}_1^\#$. Note that we recover images by solving (10) instead of (42). In fact, the two problems are equivalent. If $\mathbf{x}^\#$ is a minimizer of the (42) with some $\epsilon > 0$, there necessarily exists $\lambda \geq 0$ such that $\mathbf{x}^\#$ is a minimizer of (10) [39,40]. It should be pointed out that ϵ is related to the noise level and the optimal choice of λ is related to $\mathbf{x}^\#$ and ϵ. In this paper, we assume that the noise level is known for (42), and λ is empirically chosen in

(10) according to the noise level. In real SAR applications, the noise level can be measured when no SAR echo arrives at the radar receiver. In future works, we will consider how to deal with unknown noise level adaptively.

In summary, the above theoretical analysis shows that, using independent measurements, the proposed sampling system has smaller recovery error and enables lower sampling/data rate. This conclusion is of great importance since our goal is to achieve low-rate sampling and meanwhile provide robust recovery performance for sub-Nyquist SAR.

6. Simulations

In this section, we conduct several simulations to evaluate the proposed QuadCS-based SAR imaging performance. Results on both synthetic data and real SAR data collected by TerraSAR-X and RADARSAT-1 missions are presented. Comparisons are made between Xampling and QuadCS with equal or independent chipping sequences. The relative error, $\|\mathbf{x} - \mathbf{x}^{\#}\|_2 / \|\mathbf{x}\|_2$, and the relative root-mean-square error (RRMSE), $\mathbb{E}\left\{\|\mathbf{x} - \mathbf{x}^{\#}\|_2 / \|\mathbf{x}\|_2\right\}$, are used as the imaging performance metrics.

The simulation scheme is shown in Figure 3 and main radar parameters are summarized in Table 1. Given a synthetic image, real SAR single-look complex (SLC) image (as shown in Figure 4) or a raw data matrix, we use the operations (29) or (34) to generate sub-Nyquist measurements. Similar methods for SAR raw data generation can be found in [41,42]. To model the measurement noise, an additive white Gaussian noise (AWGN) is added. The regularization parameter and the maximum number of iterations are empirically set to 10^{-3} and 200, respectively. For visualization, the recovered images are mapped to grayscale images by the following map

$$
x_i^{gray} = \begin{cases} |x_i|/\tau, & |x_i| \le \tau; \\ 1, & \text{else,} \end{cases} \tag{44}
$$

where τ is a threshold for preventing the image from being almost black due to the presence of strong scattering points. A default value provided by a SAR image processing toolbox (code "*plotimage_sar.m*" in PPB toolbox [43]) is adopted in our simulations, which is given by the sum of mean and triple standard deviation of the amplitude image. For ease of description, we refer to QuadCS-IndSeq and QuadCS-EquSeq as QuadCS with independent and equal chipping sequences, respectively, and denote the compression ratio *M/N* by α.
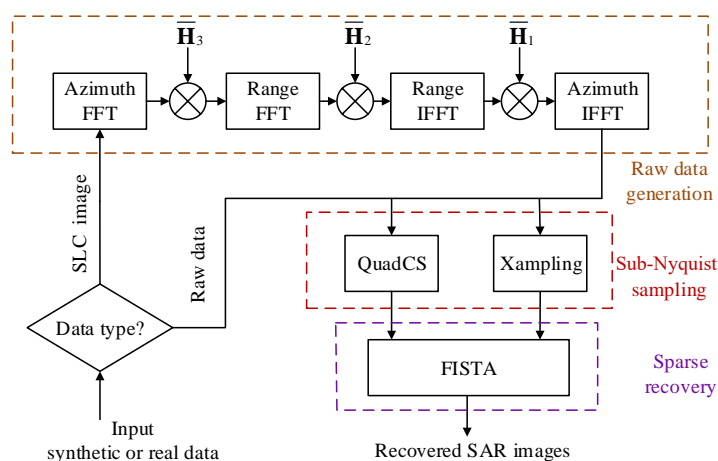

Figure 3. Block diagram of simulations. We take synthetic, real images and real raw data as input data, respectively. The simulated data output by QuadCS and Xampling are processed by a sparse recovery algorithm.

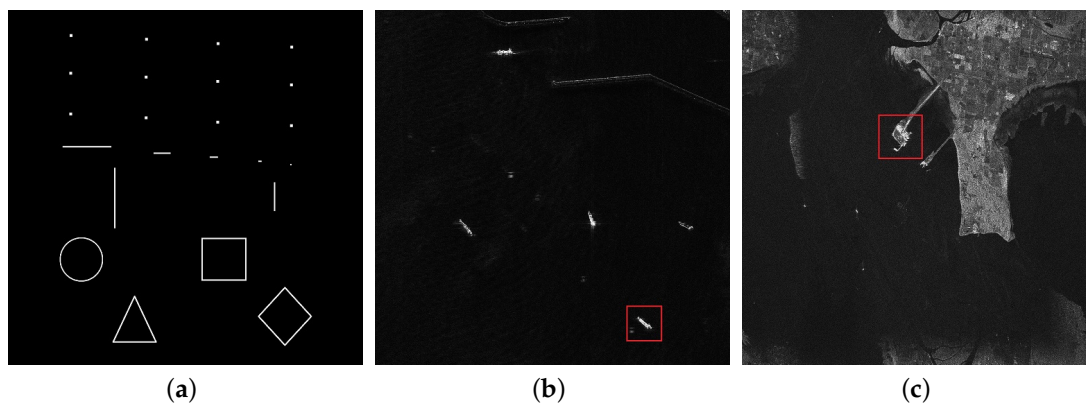| Parameter	Value
Carrier frequency	5.3 GHz
Fast-time frequency bandwidth	30.11 MHz
Pulse width	41.74 μs
Pulse repetition frequency	1256.98 Hz
Effective radar velocity	7062 m/s
Slant range of scene center	150.1 km

(**a**)   (**b**)   (**c**)

Figure 4. Scenes of interest in simulations. (**a**) a synthetic image; (**b**) a real SAR SLC image, collected by TerraSAR-X on 1 May 2008, and the imaged area is near the coast of Barcelona; (**c**) an SLC image focused from RADARSAT-1 raw data and the red-framed area is the Tsawwassen Ferry Port.

6.1. Simulations with Synthetic Data

Firstly, we use some synthetic geometries as a SLC image, as shown in Figure 4a, to evaluate the imaging performance. The compression ratio is set to 1/8 and 1/16, respectively, and the SNR equals 10 dB. As shown in the first column of Figure 5, QuadCS-IndSeq recovers these geometries very well, while QuadCS-EquSeq and Xampling cannot reconstruct horizontal lines clearly. Namely, QuadCS-EquSeq and Xampling cannot recover images in range dimension because of low sampling rate and the usage of fixed measurement matrix in each observation. At the smaller compression ratio 1/16, QuadCS-IndSeq can still recover all geometries with some performance loss, but QuadCS-EquSeq and Xampling cannot reconstruct the geometries correctly. Obviously, the proposed sampling system has the best performance.

Next, we investigate the relationship between RRMSE and the image sparsity under different compressive ratios and SNRs. We choose random subsets of the test image and set its entries to be i.i.d. random variables obeying uniform distribution on $[0, 1]$. The image recovery conducts 100 times and the resulting RRMSE curves are shown in Figure 6. All the figures show that, as the sparsity increases, the recovery errors of all sampling schemes increases. This can be explained by (4) that the best *s*-term approximation error increases as the sparsity increases. On the other hand, the figures also show that QuadCS-IndSeq has the lowest recovery error in all simulation cases. This is because QuadCS-IndSeq has better RIP, as disclosed in Theorem 2. Furthermore, QuadCS is superior to Xampling in all cases, no matter what the independent measurement strategy is utilized. The reason is that Xampling conducts frequency-domain random sampling by several bandpass filters, so that the randomness is not enough and thus the RIP of corresponding sensing matrix is not very well. Hence, QuadCS is superior to Xampling for sub-Nyqusit SAR.
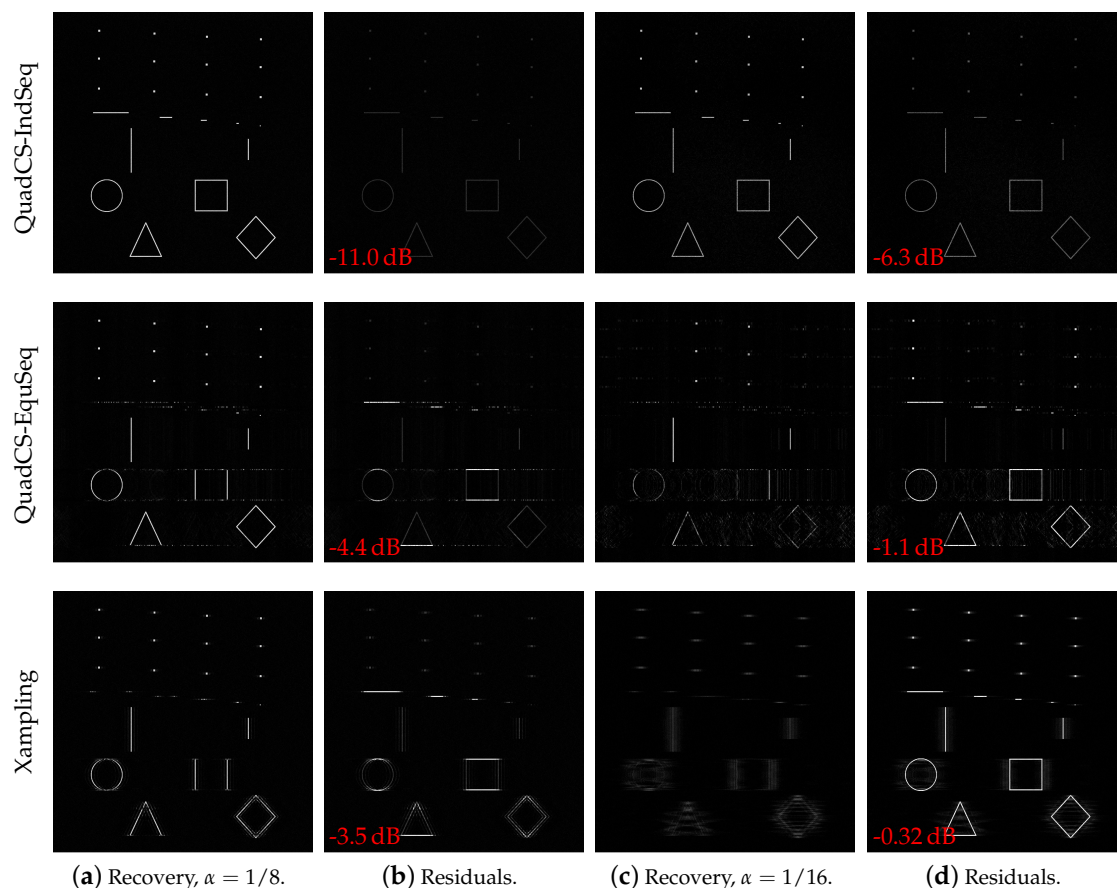
Figure 5. Recovery of the synthetic image in 10 dB AWGN. From top to bottom: QuadCS with independent chipping sequences (Quad-IndSeq), QuadCS with equal chipping sequences (Quad-EquSeq), Xampling. From left to right: recovered images and corresponding residuals with compression ratios 1/8 and 1/16. Range and azimuth are along the horizontal and vertical axes, respectively. The relative errors (measured in dB) of these images are shown at the lower left corners of their residual images, respectively.

According to each row subfigures, we find that QuadCS-IndSeq has more performance advantages in the high SNR case (SNR = 20 dB), since the RMMSE relies on the best s-term approximation error in this case. Specifically, as seen from the middle left subfigure in Figure 6, where SNR is 20 dB and compression ratio is 1/16, when the sparsity is 13‰, the RRMSE of QuadCS-IndSeq is about -14.2 dB, while those of QuadCS-EquSeq and Xampling are about -8.8 dB and -4.4 dB, respectively. On the other hand, according to each column subfigure, we know that the image recovery performance degrades faster with the lower compressive ratio, which is the inherent characteristic of sub-Nyquist sampling systems. However, QuadCS-IndSeq still has the best image recovery performance. In summary, our sampling system considerably improves the imaging performance under relatively large sparsity and low sampling rate. However, with the decrease of SNR, the recovery errors mainly come from the noise, as shown in (43); then, the improved performance is gradually diluted.
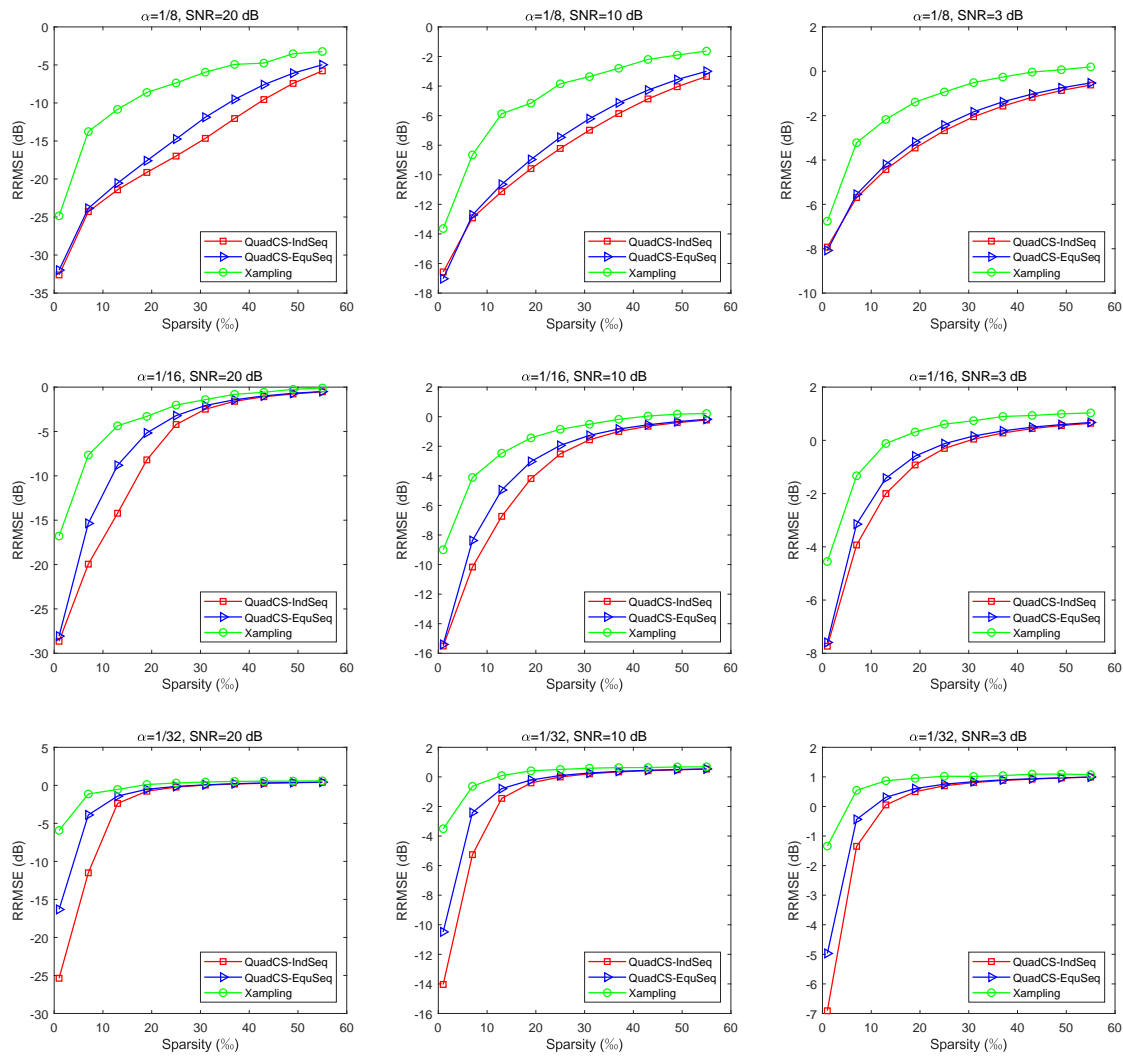
Figure 6. RRMSE vs. sparsity (averaged by 100 times). From top to bottom: compression ratios, denoted by α, are 1/8, 1/16, and 1/32, respectively. From left to right: SNR = 20 dB, SNR = 10 dB and SNR = 3 dB, respectively. We can see that QuadCS-IndSeq has the smallest RRMSE in all cases.

6.2. Simulations with Real SAR Data

Now, we use a real SLC image, shown in Figure 4b, to evaluate the proposed sampling system. The image with size 2048 × 2048 is collected by TerraSAR-X on 4 July 2008, which contains several ships near the coast of Barcelona. For a clear display, we show 200 × 200 area of the recovered images in Figure 7. We find that QuadCS-IndSeq has a smaller recovery error than QuadCS-EquSeq and Xampling.

We also use RADARSAT-1 raw data collected on 16 June 2002 to evaluate the proposed system. Imaging results using the raw data is shown in Figure 4c. Because the scene is not reasonably sparse, we consider larger settings of compression ratios, namely, 1/4 and 1/8 , respectively. Figure 8 shows a 500 × 500 area of the recovered images. As we noted in Section 1, the recovery error greatly depends on the sparsity of the image. Due to lack of sparsity, the recovery errors are large for all simulated schemes as shown in Figure 8. However, QuadCS-IndSeq still has the lowest recovery errors.
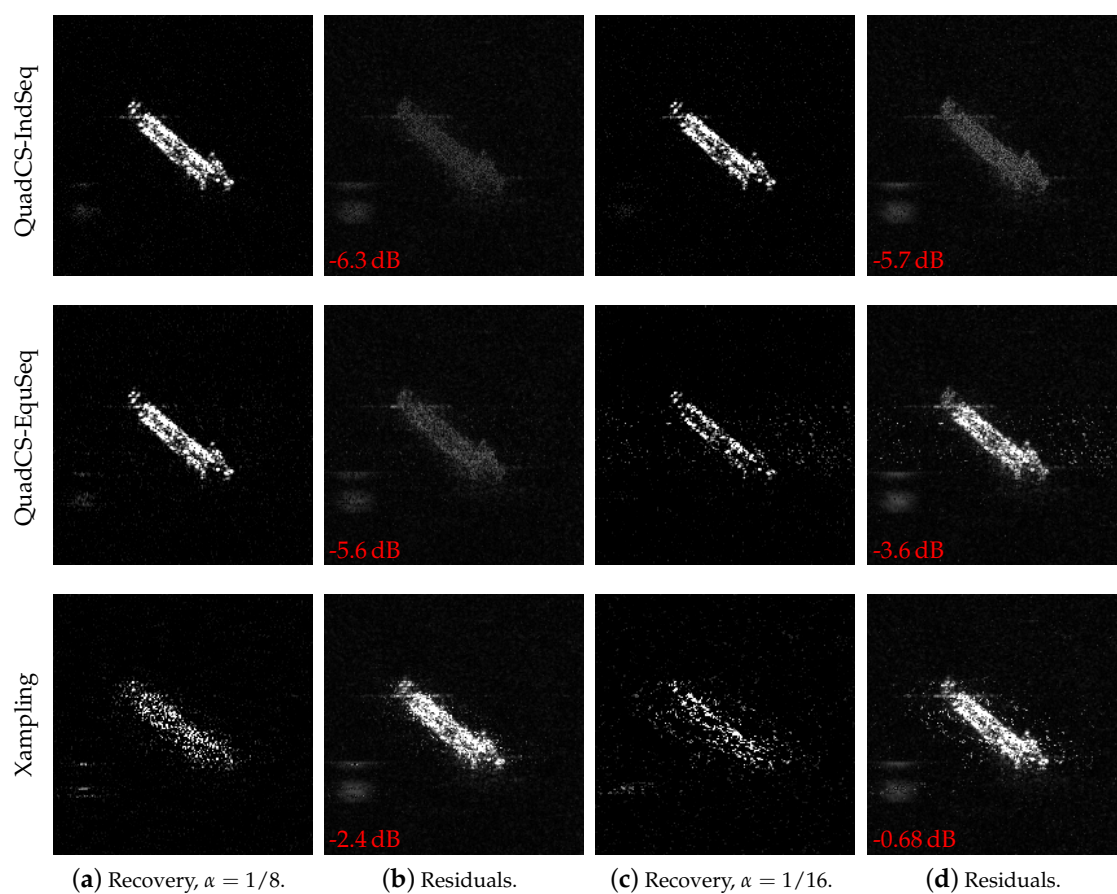
Figure 7. Recovery of the TerraSAR-X SLC image. From top to bottom: QuadCS-IndSeq, QuadCS-EquSeq, Xampling. From left to right: recovered images and corresponding residuals with compression ratios 1/8 and 1/16. Range and azimuth are along the horizontal and vertical axes, respectively. The relative errors for these images are shown at the lower left corners of their residual images, respectively. For the sparse scene, QuadCS-IndSeq achieves -5.7 dB even at the compression ratio 1/16.
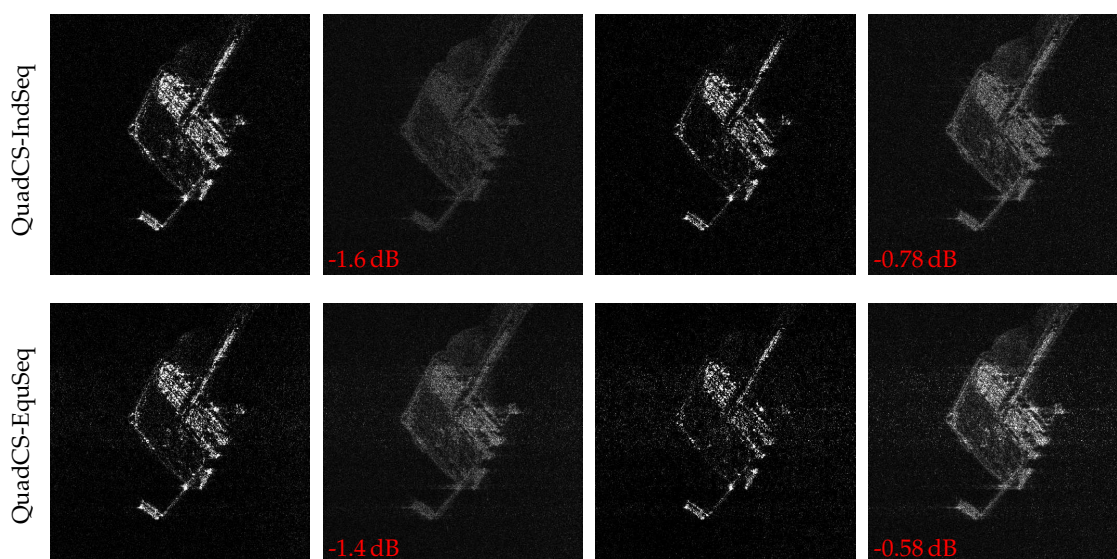

Figure 8. *Cont.*

(**a**) Recovery, $\alpha = 1/4$.   (**b**) Residuals.   (**c**) Recovery, $\alpha = 1/8$.   (**d**) Residuals.
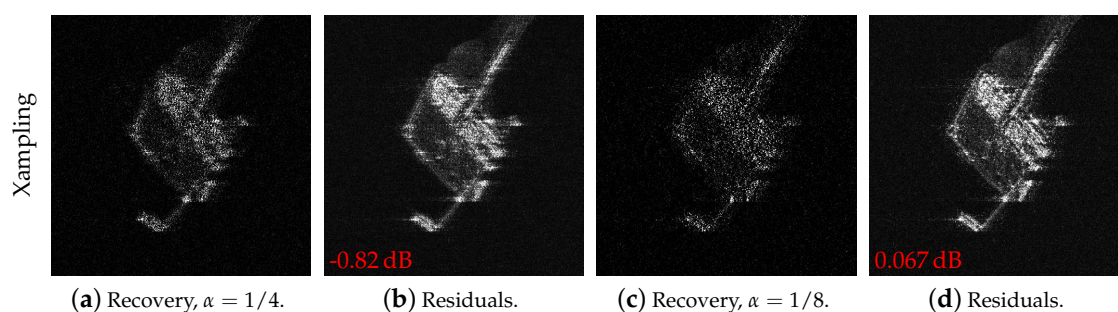
Figure 8. Recovery of the RADARSAR-1 image. From top to bottom: QuadCS-IndSeq, QuadCS-EquSeq, Xampling. From left to right: recovered images and corresponding residuals with compression ratios 1/4 and 1/8. Range and azimuth are along the horizontal and vertical axes, respectively. The relative errors for these images are shown at the lower left corners of their residual images, respectively. Due to lack of scene sparsity, all of the simulated systems cannot provide satisfactory performances.

7. Conclusions

In this paper, we presented an efficient sampling system for sub-Nyquist acquisition of SAR data. The proposed system is based on the QuadCS, which is easy to insert randomness into the sensing matrix. With independent measurements in each observation, our system has more randomness than that with fixed measurements in all observations, and thus the corresponding sensing matrix has better image recovery performance. Based on the RIP analysis, we showed that the proposed system has smaller RIC, which can lead to lower sampling/data rate and smaller recover error.

Next, we provided a fast image formation method. Through the frequency-domain representation, we found that the sensing matrix is a block and Toeplitz matrix and its matrix-vector products can be decomposed into several fast operators. In virtue of matrix decomposition, we established a first-order fast SAR imaging algorithm, whose storage and computational complexity are much less than that of the off-the-shelf recovery methods. By using both synthetic data and real data, we made comparisons among QuadCS with independent measurements, QuadCS with fixed measurements, and Xampling. Numerical results showed that the QuadCS with independent measurements achieves much better image qualities than the other simulated schemes.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic aperture radar
RIP	Restricted isometry property
CS	Compressive sensing
ADC	Analog-to-digital converter
MLFSR	Maximal-length linear feedback shift register
FFT	Fast Fourier transform
PRI	Pulse repetition interval
CSA	Chirp scaling algorithm

Appendix A

Appendix A.1

We use the tail bound on Lipschitz functions of i.i.d. random variables to derive the CoM inequality.

Firstly, we describe $\|\mathbf{A}\mathbf{x}\|_2$ as a Lipschitz function. Let us decompose the vector $\boldsymbol{\theta}_l$ in (27) as follows:

$$\boldsymbol{\theta}_l = \mathbf{F}_L \boldsymbol{\rho}_l = \mathbf{F}_L \mathbf{S}^1 \mathbf{F}_L \mathbf{p}_l = \mathbf{J}\mathbf{p}_l, \tag{A1}$$

where $\mathbf{p}_l$ is the discrete binary samples of $p_l(t)$, $\mathbf{S}^1$ circularly shifts the elements of $\mathbf{F}_L \mathbf{p}_l$ from top to bottom by one position, and $\mathbf{J} = \mathbf{F}_L \mathbf{S}^1 \mathbf{F}_L$. In (A1), the Fourier coefficient $\boldsymbol{\rho}_l$ is represented by the DFT of the length-L sequence $\mathbf{p}_l$. The decomposition (A1) is to represent $\boldsymbol{\theta}_l$ by the random Bernoulli sequence $\mathbf{p}_l$. Let us define

$$\mathbf{U} = \mathbf{F}_L^* \boldsymbol{\Delta}_a \mathbf{F}_r \mathbf{Y}, \tag{A2}$$

we can rewrite (29) as

$$\mathbf{Y}^{cs} = \sqrt{\frac{L}{M_r}} \mathbf{F}_m^* \boldsymbol{\Delta}_r \mathbf{F}_L \left(\boldsymbol{\Theta} \circ \mathbf{U} \right). \tag{A3}$$

Define a matrix

$$\boldsymbol{\Gamma} = \sqrt{\frac{L}{M_r}} \left(\mathbf{I}_{N_a} \otimes \mathbf{F}_m^* \boldsymbol{\Delta}_r \mathbf{F}_L \right) \operatorname{diag}\left(\boldsymbol{\Theta} \right), \tag{A4}$$

and denote $\mathbf{u} = \operatorname{vec}\left(\mathbf{U} \right)$; then, (A3) can be written in a matrix-vector multiplication form as

$$\mathbf{y}^{cs} = \mathbf{A}\mathbf{x} = \boldsymbol{\Gamma}\boldsymbol{u}, \tag{A5}$$

where $\mathbf{u} = \left(\mathbf{I}_{N_a} \otimes \mathbf{F}_L^* \boldsymbol{\Delta}_a \mathbf{F}_r \right) \mathbf{D}\mathbf{x}$. In the following, we replace $\|\mathbf{A}\mathbf{x}\|_2$ by $\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2$ for the convenience of proof. With the equivalence between (A3) and (A5), $\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2$ can be represented as

$$
\begin{aligned}
\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2 &= \frac{L}{M_r} \left\| \mathbf{F}_m^* \boldsymbol{\Delta}_r \mathbf{F}_L \left(\boldsymbol{\Theta} \circ \mathbf{U} \right) \right\|_F^2 \\
&= \frac{L}{M_r} \sum_l \left\| \mathbf{F}_m^* \boldsymbol{\Delta}_r \mathbf{F}_L \operatorname{diag}\left(\mathbf{u}_l \right) \boldsymbol{\theta}_l \right\|_2^2 \\
&= \frac{L}{M_r} \sum_l \left\| \mathbf{F}_m^* \boldsymbol{\Delta}_r \mathbf{F}_L \operatorname{diag}\left(\mathbf{u}_l \right) \mathbf{F}_L^* \mathbf{F}_L \mathbf{J}\mathbf{p}_l \right\|_2^2 \\
&= \frac{L}{M_r} \left\| \widehat{\boldsymbol{\Delta}} \mathbf{R} \widehat{\mathbf{F}}_L \widehat{\mathbf{J}} \mathbf{p} \right\|_2^2,
\end{aligned}
\tag{A6}
$$

where

$$\mathbf{p} = \text{vec}(\mathbf{P}), \tag{A7}$$

$$\widehat{\boldsymbol{\Delta}} = \mathbf{I}_{N_a} \otimes \boldsymbol{\Delta}_r, \tag{A8}$$

$$\widehat{\mathbf{F}}_L = \mathbf{I}_{N_a} \otimes \mathbf{F}_L, \tag{A9}$$

$$\widehat{\mathbf{J}} = \mathbf{I}_{N_a} \otimes \mathbf{J}, \tag{A10}$$

$$\mathbf{R} = \text{diag}\left(\mathbf{F}_L \text{diag}\left(\mathbf{u}_1\right) \mathbf{F}_L^*, \cdots, \mathbf{F}_L \text{diag}\left(\mathbf{u}_{N_a}\right) \mathbf{F}_L^*\right)$$
$$= \widehat{\mathbf{F}}_L \text{diag}\left(\mathbf{U}\right) \widehat{\mathbf{F}}_L^*. \tag{A11}$$

$\mathbf{P} = [\mathbf{p}_1, \ \mathbf{p}_2, \ \cdots, \ \mathbf{p}_{N_a}]$, and $\mathbf{u}_l$ is the l-th column of $\mathbf{U}$.

Let us define a function as

$$f(\mathbf{p}) = \|\boldsymbol{\Gamma}\boldsymbol{u}\|_2 = \sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\widehat{\mathbf{F}}_L\widehat{\mathbf{J}}\mathbf{p}\right\|_2. \tag{A12}$$

For a fixed vector $\mathbf{u}$, we prove that $f(\mathbf{p})$ is a Lipschitz function of the random vector $\mathbf{p}$ or its matrix form $\mathbf{P}$. Let $\mathbf{p}^{\text{I}}$ and $\mathbf{p}^{\text{II}}$ be two independent copies of $\mathbf{p}$, and we have

$$\left| f\left(\mathbf{p}^{\text{I}}\right) - f\left(\mathbf{p}^{\text{II}}\right) \right|$$
$$= \sqrt{\frac{L}{M_r}}\left| \left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\widehat{\mathbf{F}}_L\widehat{\mathbf{J}}\mathbf{p}^{\text{I}}\right\|_2 - \left\|\widehat{\boldsymbol{\Delta}}_r\mathbf{R}\widehat{\mathbf{F}}_L\widehat{\mathbf{J}}\mathbf{p}^{\text{II}}\right\|_2 \right|$$
$$\leq \sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\widehat{\mathbf{F}}_L\widehat{\mathbf{J}}\left(\mathbf{p}^{\text{I}} - \mathbf{p}^{\text{II}}\right)\right\|_2 \tag{A13}$$
$$\leq \sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\right\|_2\left\|\widehat{\mathbf{F}}_L\widehat{\mathbf{J}}\right\|_2\left\|\left(\mathbf{p}^{\text{I}} - \mathbf{p}^{\text{II}}\right)\right\|_2$$
$$= \sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\right\|_2\left\|\left(\mathbf{p}^{\text{I}} - \mathbf{p}^{\text{II}}\right)\right\|_2.$$

It can be seen that $\widehat{\boldsymbol{\Delta}}\mathbf{R}$ is a submatrix of $\mathbf{R}$, so

$$\sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}}\mathbf{R}\right\|_2 \leq \sqrt{\frac{L}{M_r}}\|\mathbf{R}\|_2 = \sqrt{\frac{L}{M_r}}\|\mathbf{u}\|_\infty := \sigma. \tag{A14}$$

Hence, $f(\mathbf{p})$ is a σ-Lipschitz function of $\mathbf{p}$.

Secondly, we use the *Talagrand concentration inequality* (Theorem 2.1.13 of [44]), which is included below for completeness, to derive the CoM inequality of $\|\mathbf{Ax}\|_2$.

Lemma A1. *Let* $\beta > 0$, *and let* $\boldsymbol{\zeta} = [\zeta_1, \cdots, \zeta_n]$ *be a variable with independent entries* $|\zeta_i| \leq \beta$ *for all* $1 \leq i \leq n$. *Let* $f : \mathbb{C}^n \to R$ *be a* σ-*Lipschitz convex function. Then, for any* δ, *one has*

$$\mathbb{P}\left\{|f(\boldsymbol{\zeta}) - \mathbb{E}f(\boldsymbol{\zeta})| \geq \delta\right\} \leq C_3 \exp\left(-\frac{C_4\delta^2}{\sigma^2\beta^2}\right), \tag{A15}$$

for some universal positive constants C_3, C_4, *where* $\mathbb{E}f(\boldsymbol{\zeta})$ *is the expectation of* $f(\boldsymbol{\zeta})$. *In [45], available values for the constants are given by* $C_3 = 4$, $C_4 = 1/16$.

In the setting of independent chipping sequences, $\mathbf{p}$ has i.i.d. ± 1 Bernoulli entries. It can be seen from (A1) that $\boldsymbol{\theta}_l$ also has i.i.d entries. Observing (A4), we find that $\mathbb{E}\{\boldsymbol{\Gamma}^*\boldsymbol{\Gamma}\} = \mathbf{I}$. Without loss of generality, let us assume $\|\mathbf{u}\|_2 = 1$. Then, we have

$$\mathbb{E}\left\{\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2\right\} = \|\mathbf{u}\|_2^2 = 1. \tag{A16}$$

According to (A15),

$$\mathbb{P}\left\{|f(\mathbf{p}) - \mathbb{E}\{f(\mathbf{p})\}| \geq \delta\right\} \leq 4\exp\left(-\frac{\delta^2}{16\sigma^2}\right). \tag{A17}$$

Since $f(\mathbf{p}) = \|\boldsymbol{\Gamma}\boldsymbol{u}\|_2$, we have

$$\mathbb{P}\left\{|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2 - 1| \geq \delta\right\} \leq 4\exp\left(-\frac{\delta^2}{16\sigma^2}\right). \tag{A18}$$

As $|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2 - 1| > \sqrt{\delta}$ is sufficient to establish $\left|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2 - 1\right| > \delta$, we have

$$\mathbb{P}\left\{\left|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2 - 1\right| \geq \delta\right\} \leq 4\exp\left(-\frac{\delta}{16\sigma^2}\right). \tag{A19}$$

Now, for an arbitrary vector $\mathbf{u}$, we have

$$\mathbb{P}\left\{\left|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2 - \|\mathbf{u}\|_2^2\right| \geq \delta \|\mathbf{u}\|_2^2\right\} \leq 4\exp\left(-\frac{\delta}{16\sigma^2}\right), \tag{A20}$$

with

$$\sigma = \sqrt{\frac{L}{M_r}} \frac{\|\mathbf{u}\|_\infty}{\|\mathbf{u}\|_2}. \tag{A21}$$

In the setting of equal chipping sequences, each sequence $\mathbf{p}_l$ has i.i.d. Bernoulli entries but $\mathbf{p}_l$, $l \in [N_a]$ are equal, i.e., $\mathbf{P}$ has identical columns,

$$\mathbf{P} = [\mathbf{p}_1, \ \mathbf{p}_1, \ \cdots, \ \mathbf{p}_1]. \tag{A22}$$

Let $\mathbf{p}_1^{\text{I}}$ and $\mathbf{p}_1^{\text{II}}$ be the first columns of two independent copies $\mathbf{P}^{\text{I}}$ and $\mathbf{P}^{\text{II}}$, respectively, and $\mathbf{p}^{\text{I}} = \text{vec}(\mathbf{P}^{\text{I}})$, $\mathbf{p}^{\text{II}} = \text{vec}(\mathbf{P}^{\text{II}})$, then we have $\|\mathbf{p}^{\text{I}} - \mathbf{p}^{\text{II}}\|_2 = \sqrt{N_a}\|\mathbf{p}_1^{\text{I}} - \mathbf{p}_1^{\text{II}}\|_2$. We therefore have

$$\begin{aligned}\left|f(\mathbf{p}^{\text{I}}) - f(\mathbf{p}^{\text{II}})\right| &\leq \sqrt{\frac{L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}\mathbf{R}}\right\|_2\left\|(\mathbf{p}^{\text{I}} - \mathbf{p}^{\text{II}})\right\|_2 \\ &= \sqrt{\frac{N_a L}{M_r}}\left\|\widehat{\boldsymbol{\Delta}\mathbf{R}}\right\|_2\left\|(\mathbf{p}_1^{\text{I}} - \mathbf{p}_1^{\text{II}})\right\|_2,\end{aligned} \tag{A23}$$

so $f(\mathbf{p})$ is a $\sqrt{N_a}\sigma$-Lipschitz function of $\mathbf{p}_1$. Following the same process from (A16) to (A20), we have

$$\mathbb{P}\left\{\left|\|\boldsymbol{\Gamma}\boldsymbol{u}\|_2^2 - \|\mathbf{u}\|_2^2\right| \geq \delta \|\mathbf{u}\|_2^2\right\} \leq 4\exp\left(-\frac{\delta}{16\sigma^2 N_a}\right). \tag{A24}$$

Next, we derive the upper bound of σ for s-sparse vectors $\mathbf{x}$. Let us define

$$\mathbf{K} = \left(\mathbf{I}_{N_a} \otimes \mathbf{F}_L^* \boldsymbol{\Delta}_a \mathbf{F}_r\right)\mathbf{D}, \tag{A25}$$

such that $\mathbf{u} = \mathbf{K}\mathbf{x}$. The upper bound of σ depends on the maximum amplitude of all entries in $\mathbf{K}$, i.e., $\max_{i,j}|(\mathbf{K})_{i,j}|$.

It is easy to know that $(\mathbf{I}_{N_a} \otimes \mathbf{F}_L^* \mathbf{\Delta}_a \mathbf{F}_r)$ in (A25) performs Fourier transform, zero-padding, and inverse Fourier transform in range dimension. According to the *principle of stationary phase* (POSP) [35], each column of $\mathbf{K}$ has the same envelope as the corresponding atom in $\mathbf{D}$. Actually, $\mathbf{K}$ is the resample of $\mathbf{D}$ at a higher rate followed by normalization. Thus, we have

$$\max_{i,j} \left| (\mathbf{K})_{i,j} \right| = \sqrt{\frac{N_r}{L}} \max_{i,j} \left| (\mathbf{D})_{i,j} \right|. \tag{A26}$$

Based on this consideration, we now determine $\max_{i,j} |(\mathbf{K})_{i,j}|$ by analyzing $\mathbf{D}$.

Let us denote the n-th column of $\mathbf{D}$ by the vector $\mathbf{d}_n$, and reshape $\mathbf{d}_n$ into an N_a-column matrix $\mathbf{D}_n$. As seen from (6), in range domain, each column of $\mathbf{D}_n$ is a LFM signal, while, in an azimuth domain, the echo data is weighted by the antenna beam pattern. Since $\mathbf{D}$ is a unitary matrix, $\|\mathbf{D}_n\|_F = 1$. Now, we determine the maximum amplitude of all entries in $\mathbf{D}_n$. With an ideal beam pattern, i.e., $w_a(\eta) = 1$, the amplitudes of all non-zero entries in $\mathbf{D}_n$ are $\sqrt{1/(\xi N)}$. In a practical case, the antenna beam pattern will bring a weight into the non-zero amplitudes as $\tilde{w}_a(lT - \eta_n)\sqrt{1/(\xi N)}$, where the weight is

$$\tilde{w}_a(lT - \eta_n) = \frac{w_a(lT - \eta_n)T_a}{\sqrt{\int_{-T_a/2}^{T_a/2} w_a^2(\eta)\, d\eta}}. \tag{A27}$$

Since the above discussion holds for all n columns, we have

$$\max_{i,j} \left| (\mathbf{D})_{i,j} \right| = \max_{n,i,j} \left| (\mathbf{D}_n)_{i,j} \right| = \frac{\omega}{\sqrt{\xi N}}, \tag{A28}$$

where $\omega = \tilde{w}_a(0)$. According to (A26) and (A28), we have

$$\max_{i,j} \left| (\mathbf{K})_{i,\,j} \right| = \omega \sqrt{\frac{N_r}{\xi NL}}. \tag{A29}$$

Let $\mathbf{x}$ be a s-sparse vector. Recalling the definition of σ in (A21), we have

$$\begin{aligned}
\sigma &= \sqrt{\frac{L}{M_r}} \frac{\|\mathbf{u}\|_\infty}{\|\mathbf{u}\|_2} = \sqrt{\frac{L}{M_r}} \frac{\|\mathbf{K}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_2} \\
&\leq \sqrt{\frac{L}{M_r}} \sup_{\|\mathbf{x}\|_2=1,\, |\text{supp}(\mathbf{x})| \geq s} \|\mathbf{K}\mathbf{x}\|_\infty \\
&\leq \sqrt{\frac{L}{M_r}} \max_{i,j} \left| (\mathbf{K})_{i,j} \right| \sqrt{s} \leq \sqrt{\frac{s}{\omega^2 \xi M}}.
\end{aligned} \tag{A30}$$

Finally, because $\|\mathbf{A}\mathbf{x}\|_2^2 = \|\mathbf{\Gamma}\mathbf{u}\|_2^2$ and $\|\mathbf{u}\|_2 = \|\mathbf{x}\|_2$, by combining (A20), (A24) and (A30), we have

$$\mathbb{P}\left(\left| \|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \geq \delta \|\mathbf{x}\|_2^2 \right) \leq 4 \exp\left(-\frac{\delta \xi \omega^2 M}{16 \chi s} \right), \tag{A31}$$

where χ denotes the two settings of chipping sequences as given in (39). The proof is completed.

Appendix A.2

Our proof below is rooted in Chapter 6 of [38]. However, some essential modifications are necessary so as to adapt the results to our case.

Firstly, we derive a uniform bound for a fixed subset. Let $S \subset [N]$ be a subset of cardinality s. According to Proposition C3 in [38] which gives a bound of covering numbers of unit ball, for $\rho \in (0, 1/2)$, there exists finite subsets $\{Z_i | i = 1, 2, \cdots, \text{card}(Z)\}$ of the unit ball

$B_S = \left\{ \mathbf{x} \in \mathbb{C}^N, \ \text{supp}(\mathbf{x}) \subset S, \ \|\mathbf{x}\|_2 \le 1 \right\}$, which satisfies $\text{card}(Z) \le (1 + 1/\rho)^{2s}$ and $\min_{\mathbf{z} \in Z} \|\mathbf{x} - \mathbf{z}\|_2 \le \rho$, for all $\mathbf{x} \in B_S$. Let us choose one element from each subset Z_i, i.e., $\mathbf{z}_i \in Z_i$ for $i = 1, 2, \cdots, \text{card}(Z)$. According to Lemma 1, the concentration inequality gives for $t \in (0, 1)$,

$$
\begin{aligned}
&\mathbb{P}\left\{ \exists \mathbf{z}_i, \ \left| \|\mathbf{A}\mathbf{z}_i\|_2^2 - \|\mathbf{z}_i\|_2^2 \right| \ge t \|\mathbf{z}_i\|_2^2 \right\} \\
&\le \sum_{i=1}^{\text{card}(Z)} \mathbb{P}\left\{ \left| \|\mathbf{A}\mathbf{z}_i\|_2^2 - \|\mathbf{z}_i\|_2^2 \right| \ge t \|\mathbf{z}_i\|_2^2 \right\} \\
&\le 4 \, \text{card}(Z) \, \exp\left(-\frac{t\xi\varpi^2 M}{16\chi s} \right) \\
&\le 4 \left(1 + \frac{2}{\rho} \right)^{2s} \exp\left(-\frac{t\xi\varpi^2 M}{16\chi s} \right).
\end{aligned}
\tag{A32}
$$

Assume that the realization of the random matrix $\mathbf{A}$ yields

$$
\left| \|\mathbf{A}\mathbf{z}_i\|_2^2 - \|\mathbf{z}_i\|_2^2 \right| \le t \|\mathbf{z}_i\|_2^2, \quad \text{for all } \mathbf{z}_i.
\tag{A33}
$$

From (A32), this occurs with probability exceeding

$$
1 - 4 \left(1 + \frac{2}{\rho} \right)^{2s} \exp\left(-\frac{t\xi\varpi^2 M}{16\chi s} \right).
\tag{A34}
$$

For a proper choice of ρ, t and δ, we prove that (A34) implies $\left| \|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \le \delta$ for all $\mathbf{x} \in B_S$, i.e., $\left\| \mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_s \right\|_2 \le \delta$. For simplicity, denote $\mathbf{Q} = \mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_s$. Now, for any $\mathbf{x} \in B_S$, there exists a vector $\mathbf{z} \in \{\mathbf{z}_i | i = 1, \cdots, \text{card}(Z)\}$ satisfying $\|\mathbf{x} - \mathbf{z}\|_2 \le \rho \le 1/2$. It then follows that

$$
\begin{aligned}
|\langle \mathbf{Q}\mathbf{x}, \mathbf{x} \rangle| &= |\langle \mathbf{Q}\mathbf{z}, \mathbf{z} \rangle + \langle \mathbf{Q}(\mathbf{x} - \mathbf{z}), \mathbf{x} + \mathbf{z} \rangle| \\
&\le |\langle \mathbf{Q}\mathbf{z}, \mathbf{z} \rangle| + |\langle \mathbf{Q}(\mathbf{x} + \mathbf{z}), \mathbf{x} - \mathbf{z} \rangle| \\
&< t + \|\mathbf{Q}\|_2 \|\mathbf{x} + \mathbf{z}\|_2 \|\mathbf{x} - \mathbf{z}\|_2 \\
&\le t + 2\rho \|\mathbf{Q}\|_2.
\end{aligned}
\tag{A35}
$$

Taking the maximum over all $\mathbf{x} \in B_S$ gives

$$
\|\mathbf{Q}\|_2 < t + 2\rho \|\mathbf{Q}\|_2,
\tag{A36}
$$

and hence

$$
\|\mathbf{Q}\|_2 \le \frac{t}{1 - 2\rho}.
\tag{A37}
$$

Let us choose $t = (1 - 2\rho)\delta$, such that $\|\mathbf{Q}\|_2 < \delta$. According to (A34), we have

$$
\begin{aligned}
&\mathbb{P}\left(\left\| \mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_s \right\|_2 \ge \delta \right) \\
&\le 4 \left(1 + \frac{2}{\rho} \right)^{2s} \exp\left(-\frac{(1 - 2\rho)\delta\xi\varpi^2 M}{16\chi s} \right).
\end{aligned}
\tag{A38}
$$

Next, we derive the RIP condition. Since δ_s can be equivalently described as $\delta_s = \sup_{S \subset [N], \ \text{card}(S) = s} \left\| \mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_s \right\|_2$ [38], taking the union bound over all $\binom{N}{s}$ subsets S gives

$$
\begin{aligned}
\mathbb{P}\left(\delta_s \geq \delta\right) \\
\leq \sum_{S \subset [N],\, \mathrm{card}(S)=s} &\mathbb{P}\left(\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_s\|_2 \geq \delta\right) \\
\leq 4\binom{N}{s}\left(1+\frac{2}{\rho}\right)^{2s} &\exp\left(-\frac{(1-2\rho)\delta\xi\varpi^2 M}{16\chi s}\right) \\
\leq 4\left(\frac{eN}{s}\right)^s \left(1+\frac{2}{\rho}\right)^{2s} &\exp\left(-\frac{(1-2\rho)\delta\xi\varpi^2 M}{16\chi s}\right).
\end{aligned} \tag{A39}
$$

Then, we have

$$
\begin{aligned}
\mathbb{P}\{\delta_s < \delta\} \\
> 1 - 4\left(\frac{eN}{s}\right)^s \left(1+\frac{2}{\rho}\right)^{2s} \exp\left(-\frac{(1-2\rho)\delta\xi\varpi^2 M}{16\chi s}\right).
\end{aligned} \tag{A40}
$$

Suppose

$$
1 - 4\left(\frac{eN}{s}\right)^s \left(1+\frac{2}{\rho}\right)^{2s} \exp\left(-\frac{(1-2\rho)\delta\xi\varpi^2 M}{16\chi s}\right) \geq 1 - \eta \tag{A41}
$$

for some η, and set $\rho = 1/4$, and then we have

$$
M \geq \frac{32\chi s}{\delta\xi\varpi^2}\left(s\ln\left(\frac{eN}{s}\right) + 4s\ln(3) + \ln\left(\frac{\eta}{4}\right)\right). \tag{A42}
$$

The proof is completed.

Appendix A.3

The choice of $\delta^{\#}=\sqrt{2}-1$ for establishing recovery error bounds in (43) refers to [46]. Under the condition in Proposition 1, M, η, ϖ, $\delta^{\#}$ and N are fixed constants. Then, according to Theorem 1, we have

$$
M = C_5 N_a \left(s_1^2\left(C_6 - \ln(s_1)\right) + C_7 s_1\right), \tag{A43}
$$

$$
M = C_5 \left(s_2^2\left(C_6 - \ln(s_2)\right) + C_7 s_2\right), \tag{A44}
$$

where $C_5 = 32(\delta^{\#})^{-1}\xi^{-1}\varpi^{-2}$, $C_6 = \ln(eN) + 4\ln(3)$ and $C_7 = \ln(\eta) - 2\ln(2)$ are constants. Dividing (A43) by (A44) gives

$$
\frac{s_1^2\left(C_6 - \ln(s_1)\right) + C_7 s_1}{s_2^2\left(C_6 - \ln(s_2)\right) + C_7 s_2} = \frac{1}{N_a}. \tag{A45}
$$

In (A45), with relatively large s, the linear terms $C_7 s_1$ and $C_7 s_2$ can be ignored, and since $\ln(s_1)$, $\ln(s_2) \ll C_6$, they also can be ignored, so we only need to keep the quadratic terms, $C_6 s_1^2$ and $C_6 s_2^2$, i.e.,

$$
\frac{s_1^2}{s_2^2} \approx \frac{1}{N_a}, \tag{A46}
$$

which gives $s_2 \approx \sqrt{N_a}s_1$. Finally, inserting this result into (4) completes the proof.

References

1. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Process Mag.* **2008**, *25*, 21–30. [CrossRef]
2. Ender, J.H.G. On compressive sensing applied to radar. *Signal Process.* **2010**, *90*, 1402–1414. [CrossRef]
3. Cetin, M.; Stojanovic I.; Onhon, O.; Varshney, K.; Samadi, S.; Karl, W.C.; Willsky A.S. Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing. *IEEE Signal Process Mag.* **2014**, *31*, 27–40. [CrossRef]

4.   Aberman, K.; Eldar, Y.C. Sub-Nyquist SAR via Fourier domain range-Doppler processing. *IEEE Trans. Geosci. Remote Sens.* **2017**, 55, 6228–6244. [CrossRef]

5.   Dong, X.; Zhang, Y. A novel compressive sensing algorithm for SAR imaging. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, 7, 708–720. [CrossRef]

6.   Kirolos, S.; Laska, J.; Wakin, M.; Duarte, M.; Baron, D.; Ragheb, T.; Massoud, Y.; Baraniuk, R. Analog-to-information conversion via random demodulation. In Proceedings of the IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software, Richardson, TX, USA, 29–30 October 2006; pp. 71–74.

7.   Duarte, M.F.; Eldar, Y.C. Structured compressed sensing: From theory to applications. *IEEE Trans. Signal Process.* **2011**, *59*, 4053–4085. [CrossRef]

8.   Laska, J.; Kirolos, S.; Massoud, Y.; Baraniuk, R.; Gilbert, A.; Iwen, M.; Strauss, M. Random sampling for analog-to-information conversion of wideband signals. In Proceedings of the IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software, Richardson, TX, USA, 29–30 October 2006; pp. 119–122.

9.   Wakin, M.; Becker, S.; Nakamura, E.; Grant, M.; Sovero, E.; Ching, D.; Yoo, J.; Romberg, J.; Emami-Neyestanak, A.; Candès, E.J. A nonuniform sampler for wideband spectrally-sparse environments. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2012**, *2*, 516–529. [CrossRef]

10.   Xi, F.; Chen, S.; Liu, Z. Quadrature compressive sampling for radar signals. *IEEE Trans. Signal Process.* **2014**, *62*, 2787–2802.

11.   Mishali, M.; Eldar, Y.C.; Dounaevsky, O.; Shoshan, E. Xampling: Analog to digital at sub-Nyquist rates. *IET Circuits Devices Syst.* **2011**, *5*, 8–20. [CrossRef]

12.   Baransky, E.; Itzhak, G.; Wagner, N.; Shmuel, I.; Shoshan, E.; Eldar, Y.C. Sub-Nyquist radar prototype: Hardware and algorithm. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 809–822. [CrossRef]

13.   Song, S.; Xu, B.; Yang, J. Ship detection in polarimetric sar images using targets' sparse property. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5706–5709.

14.   Liu, Y.; Quan, Y.; Li, J.; Zhang, L.; Xing, M. SAR imaging of multiple ships based on compressed sensing. In Proceedings of the 2009 2nd Asian-Pacific Conference on Synthetic Aperture Radar, Xi'an, China, 26–30 October 2009; pp. 112–115.

15.   Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [CrossRef]

16.   Donoho, D.L.; Elad, M. Optimally sparse representation in general (non-orthogonal) dictionaries via l1 minimization. *Proc. Nalt. Acad. Sci. USA* **2003**, *100*, 2197–2202. [CrossRef]

17.   Liu, C.; Xi, F.; Chen, S.; Zhang, Y.D.; Liu, Z. Pulse-Doppler signal processing with quadrature compressive sampling. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 1217–1230. [CrossRef]

18.   Chen, S.; Xi, F.; Liu, Z. A general sequential delay-Doppler estimation scheme for sub-Nyquist pulse-Doppler radar. *Signal Process.* **2017**, *135*, 210–217. [CrossRef]

19.   Chen, S.; Xi, F.; Liu, Z. A general and yet efficient scheme for sub-Nyquist radar processing. *Signal Process.* **2018**, *142*, 206–211. [CrossRef]

20.   Ho, K.C.; Chan, Y.T.; Inkol, R. A digital quadrature demodulation system. *IEEE Trans. Aerosp. Electron. Syst.* **1996**, *32*, 1218–1227. [CrossRef]

21.   Yang, H.; Chen, S.; Xi, F.; Liu, Z. Quadrature compressive sampling SAR imaging. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 5847–5850.

22.   Kim, S.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 606–617. [CrossRef]

23.   Nesterov, Y. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Sov. Math. Dokl.* **1983**, *27*, 372–376.

24.   Donoho, D.L.; Maleki, A.; Montanari, A. Message-passing algorithms for compressed sensing. *Proc. Nalt. Acad. Sci. USA* **2009**, *106*, 18914–18919. [CrossRef] [PubMed]

25.   Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]

26.   Becker, S.R.; Candès, E.J.; Grant, M.C. Templates for convex cone problems with applications to sparse signal recovery. *Math. Program. Comput.* **2011**, *3*, 165. [CrossRef]

27. Bhattacharya, S.; Blumensath, T.; Mulgrew, B.; Davies, M. Fast encoding of synthetic aperture radar raw data using compressed sensing. In Proceedings of the IEEE/SP 14th Workshop on Statistical Signal Processing, Madison, WI, USA, 26–29 August 2007; pp. 448–452.

28. Alonso, M.T.; Lopez-Dekker, P.; Mallorqui, J.J. A novel strategy for radar imaging based on compressive sensing. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4285–4295. [CrossRef]

29. Chen, S.; Xi, F. Quadrature compressive sampling for multiband radar echo signals. *IEEE Access.* **2017**, *5*, 19742–19760. [CrossRef]

30. Michael, N.; Shah, S.; Das, J.; Sandeep, M.M.; Vijaykumar, C. Non uniform digitizer for alias-free sampling of wideband analog signals. In Proceedings of the In TENCON 2007—2007 IEEE Region 10 Conference, Taipei, Taiwan, 30 October–2 November 2007; pp. 1–4.

31. Baraniuk, R.; Davenport, M.; Duarte, M.; Hegde, C. An Introduction to Compressive Sensing; Connexions e-Textbook, 2011. Available online: http://cnx.org/content/col11133/1.5 (accessed on 14 May 2015)

32. Beck, A.; Teboulle, M. Guide to the Matlab Code for Wavelet-Based Deblurring with FISTA. Available online: http://iew3.technion.ac.il/~becka/papers/wavelet_FISTA.zip (accessed on 19 June 2018).

33. Wright, S.J.; Nowak, R.D.; Figueiredo, M.A.T. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* **2009**, *57*, 2479–2493. [CrossRef]

34. Fang, J.; Xu, Z.; Zhang, B.; Hong, W.; Wu, Y. Fast compressed sensing SAR imaging based on approximated observation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 352–363. [CrossRef]

35. Cumming, I.G.; Wong, F.H. *Digital Signal Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*; Artech House: Norwood, MA, USA, 2004.

36. Ledoux, M. *The Concentration of Measure Phenomenon*; American Mathematical Society: Washington, DC, USA, 2001.

37. Sanandaji, B.M.; Vincent, T.L.; Wakin, M.B. Concentration of measure inequalities for Toeplitz matrices with applications. *IEEE Trans. Signal Process.* **2013**, *61*, 109–117. [CrossRef]

38. Foucart, S.; Rauhut, H. *A Mathematical Introduction to Compressive Sensing*; Springer: New York, USA, 2013.

39. Figueiredo, M.A.T.; Nowak, R.D.; Wright, S.J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 586–597. [CrossRef]

40. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.

41. Khwaja, A.S.; Ferro-Famil, L.; Pottier, E. Efficient SAR raw data generation for anisotropic urban scenes based on inverse processing. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 757–761. [CrossRef]

42. Qiu, X.; Hu, D.; Zhou, L.; Ding, C. A bistatic SAR raw data simulator based on inverse ω-k algorithm. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1540–1547.

43. Deledalle, C. Probabilistic Patch-Based Filter (ppb). Available online: http://www.ceremade.dauphine.fr/~deledall/ppb.php (accessed on 16 June 2018).

44. Tao, T. *Topics in Random Matrix Theory*; American Mathematical Society: Washington, DC, USA, 2012.

45. Massart, P. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **2000**, *28*, 863–884. [CrossRef]

46. Candès, E.J. The restricted isometry property and its implications for compressed sensing. *C. R. Math.* **2008**, *346*, 589–592. [CrossRef]