

Physics-enhanced PCA for Data Compression in Edge Devices

Qianyi Chen, Jiannong Cao, *Fellow, IEEE*, and Yong Xia

Abstract—In smart cities, tremendous data are generated with edge devices continuously for scientific applications, such as structural health monitoring (SHM), leading to a high bandwidth burden to edge devices. Data compression is a typical technique for reducing data size and improving transmission efficiency. However, their operations for distinguishing redundant data might be unapplicable for a domain-specific scientific analysis and distort physical quantities of raw data (i.e., mode shapes of vibration data). Besides, they are too compute-intensive to be executed in resource-constraint edge devices. In this paper, we leverage physical knowledge to enhance a lightweight data compression method for edge devices in maintaining physical quantities. In particular, we propose physics-enhanced PCA for compressing data in a dynamic system — compressing the vibration data of a structure in the context of SHM. Physical knowledge is identified from a structure and guides the compression process to preserve the mode shape, an significant physical quantity for structures. We have formally analyzed the effectiveness of our approach and performed experiments to show that physical knowledge is essential for preserving mode shapes. Concretely, experiments in numerical and real-world structures show that physics-enhanced PCA can improve the accuracy by up to 56% compared with alternative baseline.

Index Terms—Data compression, Edge computing, Smart city, Structural health monitoring.

I. INTRODUCTION

BENEFITED from the decreased cost of sensing devices and advances in the internet of things (IoT) technology, increasing quantities of sensors with high sampling rates are deployed in urban areas, generating tremendous IoT data continuously for smart city applications [1]. Cloud computing is extensively used in these applications owing to its elastic and economical computing resources [2], [3]. Meanwhile, edge computing is attracting increasing attention as it can provide services with lower latency and higher security. As Figure 1 shows, both cloud computing and edge computing are crucial for smart city applications. The data transmission efficiency between them will determine the overall performance. Nevertheless, it becomes the bottleneck because of the limited bandwidth and unstable network in edge devices. Therefore, various Data compression methods [4]–[7] have been deployed

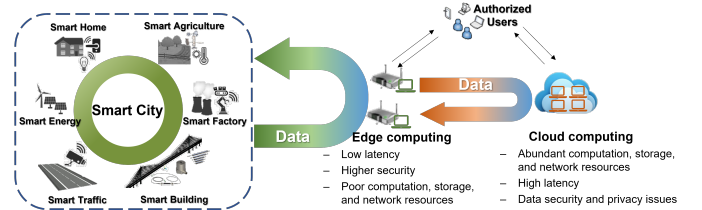


Fig. 1. Cloud computing and edge computing for Smart City Applications

on edge devices to shrink the data size and improve the transmission efficiency. Basically, they are classified into lossless data compression methods and lossy data compression methods. Lossless methods usually achieve a lower compression ratio (CR) as they are required to reconstruct the raw data perfectly. Also, their performance in compressing IoT data is very limited due to the almost random nature of such data [8], [9]. Therefore, this paper only focuses on lossy data compression methods, where redundant data are eliminated during data compression, and higher CR can be achieved.

Most lossy data compression methods distinguish the primary and redundant parts of data based on specific assumptions. For example, they assume that redundant data have smaller co-variance in a transformed coordinate system [10], smaller amplitude in the frequency domain [11], or smaller amplitude in the spatio-frequency domain [11]. Data compression is then achieved by removing those redundant parts from raw data. Nevertheless, those operations might not be applicable for some domain-specific scientific analysis, because important physical quantities contained in raw data might be eliminated or distorted during data compression. For instance in structural health monitoring (SHM) [12], researchers extract model properties such as natural frequency, mode shape, and damping ratio from the structures' dynamic responses to estimate structure's overall mechanical properties. Those model properties might be skewed during data compression, which mislead the decision making in SHM. [9], [13] have proposed physically motivated metrics for evaluating the performance in maintaining critical physical quantities. However, data compression methods themselves haven't been modified accordingly.

In order to preserve physical quantities precisely, the algorithm of a scientific analysis must be integrated with a data compression method. Nevertheless, executing the whole scientific analysis will bring much additional workload to edge devices, increasing power consumption and slowing down the

Q. Chen is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: 19074424r@connect.polyu.hk)

J. Cao is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: csjcao@comp.polyu.edu.hk)

Y. Xia is with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: ceysxia@polyu.edu.hk)

Manuscript received August 15, 2021; revised January 2, 2022.

0000-0000/00\$00.00 © 2021 IEEE

data compression process severally. To solve this challenge, we innovatively leverage physical knowledge to enhance the data compression methods. We choose SHM as the target scientific application and design a physics-enhanced principal component analysis (PCA) for compressing vibration data in edge devices. Raw vibration data are rescaled with the normalized mass matrix identified from previous measurement data. We have proved that there exists a one-to-one mapping between mode shapes and principal components extracted from rescaled vibration data. So, eliminating the component with small variance only leads to the elimination of the corresponding mode shape, the other mode shapes which dominate the vibration will not be degraded. Modal responses represent the component described by a mode shape in vibration data. Consequently, the modal parameters are accurately preserved.

The rest of this paper is organized as follow: related works are introduced in section II. In section III, a brief introduction on PCA, dynamic system simulation, and their relationship is presented. Then, history PCA and physics-enhanced PCA are introduced in section IV. A numerical experiment is introduced in section V to prove the performance of the proposed method. In section VI, an edge computing based structural health monitoring system is designed for a real structure. The proposed data compression algorithm is deployed on edge devices, and its performance with real-world vibration data is analyzed. Finally, the conclusions are drawn, and future works are listed in Section VII.

II. RELATED WORK

Existing lossy data compression methods for IoT data are based on discrete Fourier transform (DFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), compressive sampling (CS), and machine learning (ML). DFT and DCT transform data into the frequency domain, and DWT transforms data into spatio-frequency domain [11]. Data compression is then achieved by eliminating coefficients with relatively small values in the transformed domain. DFT, DCT, DWT, and their variants have been extensively used in compressing climate data [14], electrocardiogram data [15], structural vibration data [16], [17], and even power system signals [18]. As an alternative theory to Nyquist/Shannon sampling theory, CS can reconstruct signals with much fewer measurements under the assumption that signals are sparse or compressible in a certain domain such as frequency domain, spatial domain, DCT domain, wavelet domain, etc. Bao et al. [19] used compressive sampling for the first time in compressing real acceleration data collected from structures. But the CR achieved using CS is low because acceleration data are not naturally sparse in the chosen bases. Recently, ML-based data compression methods have attracted a lot of attention due to their automatic feature extraction ability and high CR. Ni et al. [12] used autoencoder to compress and recover the acceleration data with high precision under a low CR. A PCA-based data compression method is proposed by Alessio [20]. Based on the experiment on a real highway bridge, the feasibility of the proposed method is proved.

The time and space complexity of above methods in compressing an $N \times T$ matrix are compared in Table I. Except

for PCA, ML-based methods are not included in this table because their time and space complexities vary with their architectures. We assume that N is far less than T because the number of sensors is usually limited, while the number of observations for each sensor is much larger. As Table I shows, history PCA can compress the raw data with minimal time and space complexity. So, it is suitable for resource-constraint edge devices. Besides, PCA-based methods have been proved to achieve higher CR than other data compression methods [20]. Consequently, physics-enhanced PCA is designed on the basis of history PCA, and a detailed introduction is included in section 3.

TABLE I
COMPARISON OF EXISTING DATA COMPRESSION METHODS

Data compression method	Time complexity	Space complexity	ref
DFT, DCT, DWT	$O(NT \log_2 T)$	$O(NT)$	[11]
CS	$O(NT^2/CR)$	$O(NT/CR)$	[21]
PCA	$O(N^2T)$	$O(N^2)$	[22]
Distributed PCA (DRO)	$O(NT)$	$O(\sum_{i=1}^N n_i^2)$	[10]
History PCA	$O(N^3/CR^2)$	$O(N^2/CR + NB)$	[20]

Note: Distributed PCA (DRO) represents the distributed PCA with distributed power method for distributed row observation (DRO) data; N denotes the dimension of each sample; T denotes the number of observation for each sample; CR denotes compression ratio; B denotes the number of observation in each block in history PCA

III. PCA AND DYNAMIC SYSTEM SIMULATION

Principal component analysis is widely used in signal processing and data mining for data compression and feature extraction. System identification is a typical scientific analysis used to identify the modal parameters of a dynamic system, such as a structure. In this chapter, the basic concepts of PCA and dynamical system identification are introduced. PCA and its interpretation based on system identification are presented as well.

A. Principal Component Analysis

PCA, also called proper orthogonal decomposition (POD), is one of the most widely adopted feature extraction algorithms in feature engineering and data analytics. Owing to its eye-attracting performance in feature extraction, researchers usually utilize it jointly with deep learning algorithms for data compression and prevent overfitting. Also, because of its ability to extract features, PCA is also used for data denoising and data compression. It is designed to transform a set of stationary zero mean time series data into the multiplication of a linear transformation matrix and a set of stationary time series data with descending variance. The transformation process is shown in Eq. (1).

$$\mathbf{O} = \mathbf{U}\mathbf{Z} \quad (1)$$

, where $\mathbf{O} = [\mathbf{o}(t_1), \mathbf{o}(t_2), \dots, \mathbf{o}(t_i), \dots, \mathbf{o}(t_T)]$, $\mathbf{O} \in \mathbb{R}^{N \times T}$ denotes the original data matrix. $\mathbf{o}(t_i) \in \mathbb{R}^{N \times 1}$ denotes a vector formed by N observations at time t_i . $\mathbf{Z} \in \mathbb{R}^{N \times T}$ and $\mathbf{U}$ denotes the linear transformation matrix which are computed through performing singular value decomposition (SVD) on the covariance matrix $\mathbf{O}\mathbf{O}^T$. Either two columns of this matrix are orthonormal.

The PCA-based data compression is performed by removing the last several columns of $\mathbf{U}$ and corresponding rows of $\mathbf{Z}$ based on a predefined threshold for the cumulative energy. The cumulative energy can be obtained in Eq. (2).

$$C_e = \frac{\sum_{i=1}^L \sigma_i}{E} \geq \text{thre}, E = \sum_{i=1}^N \sigma_i \quad (2)$$

, where C_e denotes the cumulative energy. σ_i denotes the variance of i th row in σ_i and thre denotes the predefined threshold.

B. Dynamic System Identification

Dynamical systems [23] such as aerospace vehicles, power plants, chemical processes, control systems, and physiological systems can be described with differential equation models and state-space models. In dynamic system simulation for structures, the governing equation of motion that follows Newton's second law of motion is most frequently used to simulate the vibration of structures. Finite element method (FEM) is widely used to transform the structure into discretized spatial domain forms, after which the motion of the structure can be represented with equation (3).

$$\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \mathbf{F}(t) \quad (3)$$

, where $\mathbf{M} \in \mathbb{R}^{N \times N}$, $\mathbf{C} \in \mathbb{R}^{N \times N}$, and $\mathbf{K} \in \mathbb{R}^{N \times N}$ denote mass matrix, damping matrix, and stiffness matrix respectively for the discretized structure with N degrees of freedom. $\mathbf{q}(t) \in \mathbb{R}^{N \times 1}$ denotes the displace vector and $\mathbf{F}(t) \in \mathbb{R}^{N \times 1}$ denotes the external load vector. Except for the equation of motion, state-space model is also used in the system identification for structures. It describes the input (excitation) and output (observation) of the linear time-invariant system as shown in equation (4).

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{aligned} \quad (4)$$

, where the $\mathbf{x}(t)$ denotes the state vector, $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the input vector and output vector of the system. Matrix $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are used to describe the properties of the system. Through defining

$$\begin{aligned} \mathbf{x}(t) &= [\mathbf{q}(t)^T, \dot{\mathbf{q}}(t)^T]^T \\ \mathbf{y}(t) &= \mathbf{C}_a \ddot{\mathbf{q}}(t) + \mathbf{C}_v \dot{\mathbf{q}}(t) + \mathbf{C}_s \mathbf{q}(t) \end{aligned} \quad (5)$$

and rearranging the equations, equation (3) can be transformed into state-space model and the system matrix can be represented by mass, damping and stiffness metrics:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{M}^{-1}\mathbf{B} \end{bmatrix} \quad (6)$$

$$\mathbf{C} = [\mathbf{C}_s - \mathbf{C}_a \mathbf{M}^{-1} \mathbf{K} \quad \mathbf{C}_v - \mathbf{C}_a \mathbf{M}^{-1} \mathbf{C}] \quad (7)$$

$$\mathbf{D} = \mathbf{C}_a \mathbf{M}^{-1} \mathbf{B} \quad (8)$$

As shown above, the state-space model is a first-order differential equation, while the governing equation of motion is a second-order equation. Nevertheless, they are two models that describe an identical system in different forms. The process of identifying model parameters from measurement

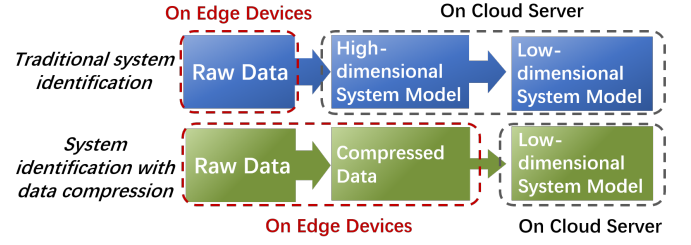


Fig. 2. The function of data compression in system monitoring

data of a dynamic system is called system identification. Most existing system identification algorithms assume real-world structures to be linear time-invariant systems during the normal operation period. Firstly, state-space models are used to fit the measurement data. After coordinate transformation, the identified state-space model is then changed into the form of motion equation, through which the physical quantities including mass, stiffness, and damping of the system can be finally obtained [24].

During system identification, the identified high-dimensional system is usually reduced to a low-dimensional one. Then, the computation can be reduced, but the most relevant mechanisms are still captured. Widely-used model reduction methods include balanced truncation [25], Galerkin projection [26], eigensystem realization algorithm (ERA) [27], balanced proper orthogonal decomposition [28] and stochastic subspace identification [29]. As Figure 2 shows, system models are finally compressed in the traditional system identification process, and there is no need to transmit raw data to the cloud. In this paper, data are compressed on edge devices before transmission. By choosing suitable CR, the loss in system identification accuracy can be minimized, but the data transmission efficiency can be dramatically improved.

C. Interpretation of PCA Based on System Identification

When the coordinates of the displacement vector $\mathbf{q}(t)$ in Eq.(3) are changed as shown in Eq. (9), the system parameters will change accordingly.

$$\mathbf{q}(t) = \mathbf{\Psi} \mathbf{a}(t) \quad (9)$$

, where $\mathbf{a}(t)$ denotes the displacement vector at time t in the transformed coordinate system. Each column of matrix $\mathbf{\Psi}$ denotes a basis of the coordinate system. Theoretically, there are infinite sets of solutions for modal parameters because the number of available coordinate systems is infinite. Among them, there is one coordinate system in which the vibration data of different freedoms can be decoupled, and this system is modal coordinate. Bases of a modal coordinate system are mode shapes, which are important physical quantities in a dynamic system. Similar to Eq. (9), the essence of PCA is also coordinate transformation. But, the coordinate transformation in PCA follows a specific rule, and the obtained coordinate transformation matrix owns specific properties. The first column of the matrix represents the direction in which the projected data have the maximum variance. The following

columns of the matrix are arranged by descending order of variance of the projection.

Because the vibration data measured by sensors are discrete in the time domain, Eq. (9) is written in the matrix form:

$$\mathbf{X} = \Phi \mathbf{A} \quad (10)$$

, Where $\mathbf{X} = [\mathbf{q}(t_1), \mathbf{q}(t_2), \dots, \mathbf{q}(t_T)]$, $\mathbf{X} \in \mathbb{R}^{N \times T}$ is the matrix formed by N sets of displacement/velocity/acceleration time series vibration data. Each set of vibration data have T samples with constant sampling frequency. Φ is the modal coordinates, and $\mathbf{A} \in \mathbb{R}^{N \times T}$ denotes the modal response in harmonic function form. When the time period is long enough, the modal responses for every two different modes are independent, which can be proved based on the orthogonality of any two Fourier basis with different frequencies [30].

Despite the similarity of Eq. (1) and (10), it has been proved by Han et al. [31] that there is no one-to-one mapping between the normal modes of a structure and proper orthogonal modes unless the mass matrix of the structure is the constant multiple of an identity matrix. Such a requirement can hardly be satisfied for real-world structures. Therefore, PCA cannot extract normal modes directly from the vibration data. Wang. et al [32] also pointed out that PCA will lead to the loses of amplitude information of mode shapes. Through comparing Eq. (9) and Eq. (10), we found that PCA can be interpreted as modal coordinates transformation if the transformation matrix in the PCA is an orthogonal matrix. Based on [31], the modal matrix becomes diagonal only when the mass matrix is the constant multiple of the identity matrix. Apparently, we cannot assume that the mass of each element to be identical. But, the normal mode shapes of a structure have the orthogonality defined as [33], [34]:

$$\Phi^T \mathbf{M} \Phi = \Delta^2 \quad (11)$$

, where Φ is the arbitrary scaled modal matrix with each column as a mode shape, $\mathbf{M}$ is the mass matrix, and Δ is the scaling factor matrix. Both $\mathbf{M}$ and Δ are diagonal matrices with positive elements. Then, the equation can be rewritten to be the multiplication of an orthogonal matrix and its transpose

$$\mathbf{S}^T \mathbf{S} = \mathbf{I} \quad (12)$$

, where $\mathbf{S} = \mathbf{M}^{\frac{1}{2}} \Phi \Delta^{-1}$. Through premultiply the vibration data with $\mathbf{M}^{\frac{1}{2}}$, Eq. (10) is transformed into

$$\mathbf{M}^{\frac{1}{2}} \mathbf{X} = \mathbf{S} \tilde{\mathbf{A}} \quad (13)$$

, where $\tilde{\mathbf{A}} = \Delta \mathbf{A}$. Row vectors in $\tilde{\mathbf{A}}$ have been rearranged in decreasing order of the variance of each row vector from the first to the last row.

As shown in Eq. (13), after premultiplied by $\mathbf{M}^{\frac{1}{2}}$, the vibration data matrix can be decomposed into the product of an orthogonal matrix and a matrix with independent row vectors. Because the ambient vibration data of a structure are usually assumed to follow the Gaussian distribution. $\mathbf{M}^{\frac{1}{2}} \mathbf{X}$ can also be transformed into the multiplication of a linear transformation matrix $\mathbf{U}$ and matrix $\mathbf{Z}$ with independent row vectors [35] following Eq. (1). To prove that there exists a one-to-one mapping between $\mathbf{U}$ and $\mathbf{S}$, we assume that:

$$\mathbf{M}^{\frac{1}{2}} \mathbf{X} = \mathbf{S}_a \tilde{\mathbf{A}}_a \quad (14)$$

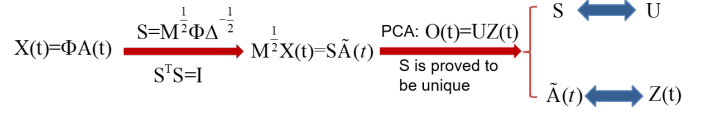


Fig. 3. Derivation for the mapping between matrix $\mathbf{S}$ and $\mathbf{U}$

, where $\mathbf{S}_a$ is an orthogonal matrix and $\mathbf{S}_a \neq \mathbf{S}$. $\tilde{\mathbf{A}}_a$ is a matrix with independent row vectors. Same as $\tilde{\mathbf{A}}$, the variances of it's row vectors decrease from the first to the last row. Substitute Eq. (13) into Eq. (14), we can obtain

$$\tilde{\mathbf{A}}_a = \mathbf{S}_a^T \tilde{\mathbf{A}} \quad (15)$$

Because $\mathbf{S}_a \neq \mathbf{S}$, $\mathbf{S}_a^T \mathbf{S}$ can't be a diagonal matrix. And there is at least one nonzero value in this matrix apart from the diagonal elements. Based on Eq. (15), each row vector in $\tilde{\mathbf{A}}_a$ is the linear combination of all row vectors in $\tilde{\mathbf{A}}$. If $\mathbf{S}_a^T \mathbf{S}$ is not a diagonal matrix. Row vectors in $\tilde{\mathbf{A}}_a$ cannot be independent, which is contradictory with the assumption. Therefore, for a given $\mathbf{M}^{\frac{1}{2}} \mathbf{X}$, $\mathbf{S}$ is unique. And there is an one-to-one mapping between $\mathbf{U}$ and $\mathbf{S}$. As has been proved that the solution for the linear transformation matrix $\mathbf{U}$ in PCA is unique. The matrix $\mathbf{U}$ computed in PCA is equal to $\mathbf{S}$. Also, $\mathbf{Z}$ is equal to $\tilde{\mathbf{A}}$, which is the product of scaling factor matrix Δ and modal responses $\mathbf{A}$. The derivation procedure has been illustrated with Figure 3

As has been proved above, there is a one-to-one mapping between the principal components extracted from $\mathbf{M}^{\frac{1}{2}}$ rescaled vibration data and mode shapes. When components with relatively small variance are removed during data compression with PCA, other mode shapes will not be degraded. Consequently, the mode shapes that dominate the vibration data are precisely preserved.

IV. PHYSICS-ENHANCED PCA

A. Normalized Mass Matrix

Various approaches [36] [37] [38] have been proposed to obtain the mass matrix or modal mass matrix from the vibration data. Identifying the first-order state-space model is usually the first step. Then the identified state-space model is transformed into the second-order motion equation, from which the physical quantities can be identified. However, identifying the accurate parameters for a state-space model is difficult because of the inevitable measurement noises, limited observations, and nonlinearity of the structures. Also, retrieving the normal modal parameters from the identified complex modes in the state-space model is hard because of the nonuniqueness issues [24] [39], which increases the difficulties in obtaining the mass matrix and modal mass matrix. Therefore, a normalized mass matrix $\mathbf{M}_n$ is used as the alternative in this paper. Because the mass matrix's orthogonality to modal matrix remains unchanged when it is normalized. Eq. (11) still holds for $\mathbf{M}_n$ as follow:

$$\Phi^T \mathbf{M}_n \Phi = \Delta_n^2 \quad (16)$$

, where Δ_n is also a diagonal matrix. $\mathbf{M}_n$ is then used to rescale the vibration data for preserving the modal properties

when using PCA. According to Eq. (11), when the modal matrix is available, the normalized mass matrix can be obtained through optimizing the non-diagonal elements in Δ_n to be zero. Δ_n can be written in elementwise as follow:

$$\begin{bmatrix} \sum_{k=1}^N \mathbf{m}_k \Phi_{k1} \Phi_{k1} & \dots & \sum_{k=1}^N \mathbf{m}_k \Phi_{k1} \Phi_{kn} \\ \dots & \sum_{k=1}^N \mathbf{m}_k \Phi_{ki} \Phi_{kj} & \dots \\ \sum_{k=1}^N \mathbf{m}_k \Phi_{kn} \Phi_{k1} & \dots & \sum_{k=1}^N \mathbf{m}_k \Phi_{kn} \Phi_{kn} \end{bmatrix} \quad (17)$$

$$\mathbf{M}_n = \text{diag}(\mathbf{m}) \quad (18)$$

, where $\mathbf{m}_k$ is the k th element in vector $\mathbf{m}$. Φ_{ij} denotes the element in i th row and j th column of the modal matrix. To compute $\mathbf{M}_n$, we defined a objective function and form the quadratic optimization problem as:

$$\begin{aligned} \min \sum_{i=1}^N \sum_{j=i+1}^N [\sum_{k=1}^N (\mathbf{m}_k \Phi_{ki} \Phi_{kj})]^2 \\ \text{s.t. } \mathbf{m}_k > 0, k = 1 : N \\ \mathbf{m}_r = 1, 1 \leq r \leq N \end{aligned} \quad (19)$$

, where $\mathbf{m}_r$ is a random element in $\mathbf{m}$. $\mathbf{m}_r$ is fixed to 1 in order to normalize the solution of this problem. And that's why $\mathbf{M}_n$ is named the normalized mass matrix. Because a structure's mass matrix can be regarded as constant in a short-time. The normalized mass matrix $\mathbf{M}_n$ also stays unchanged and we don't need to update it frequently. However, the regular update in $\mathbf{M}_n$ is needed in case of the change of mass distribution results from structural degradation or sudden disaster. And the frequency of updates should be decided according to the situation.

B. History PCA and Physics-enhanced PCA

In traditional PCA, the sample covariance matrix is computed when the whole data matrix is ready. It proposes a high requirement to edge devices on the memory and computing resources, especially when multi-channel data with high sampling frequency need to be compressed. In addition, the computation complexity of PCA increased exponentially with the increase of the dimension of input data. Distributed PCA is proposed to solve these issues [10] and streaming PCA is a type of distributed PCA that works in a dataset with continuously generated samples. By computing the principal components and eigenvectors in a streaming setting with much less memory, the computation burden of PCA is mitigated remarkably. In streaming PCA, Oja's algorithm [40] is used to transform the original problem into an optimization problem which can be solved with the stochastic gradient descent (SGD) method. Then, the PCA problem can be solved iteratively with streaming data. Different from other streaming PCA algorithms, history PCA [41] can make better use of the previous data and converges much faster. In history PCA, the contribution of historical data to the sample covariance matrix is exploited, and it is proved to converge faster than other variants of PCA. Based on the history PCA algorithm, we rescale the input data stream with the root of the normalized mass

matrix to preserve the mode shape information. Workflow of the proposed method is introduced in Alg. 1.

Algorithm 1 Physics-enhanced PCA

Input: $\mathbf{M}_n^{-\frac{1}{2}}$, k , m , and $\mathbf{X} = [\mathbf{X}_1^\top, \mathbf{X}_2^\top, \dots, \mathbf{X}_n^\top]$, $\mathbf{X}_i \in \mathbf{R}^{N \times B}$ ($1 \leq i \leq n$). $\mathbf{M}_n^{\frac{1}{2}}$ is computed through computing the root of each element in $\mathbf{M}_n$ and $\mathbf{M}_n^{-\frac{1}{2}}$ denotes the inverse of $\mathbf{M}_n^{\frac{1}{2}}$. B is the number of observations in one block. Each block is a small set of vibration data collected in a small period, and n is the number of blocks. k is the number of principal components reserved in PCA. m denotes the number of iteration in each block

Output: Compression matrix $\mathbf{Q}_\tau$

$\mathbf{H}^i \sim N(0, I_{d \times d})$, $1 \leq i \leq k$

$\mathbf{H} = \mathbf{Q}_1 \mathbf{R}_1$ (QR-decomposition)

$\tilde{\mathbf{X}}_1 = \mathbf{M}_n^{\frac{1}{2}} \mathbf{X}_1$

for $i = 1 \rightarrow m$ **do**

$\mathbf{S}_1 \leftarrow \mathbf{Q}_1 + \frac{1}{B} \tilde{\mathbf{X}}_1^\top \tilde{\mathbf{X}}_1 \mathbf{Q}_1$

$\mathbf{S}_1 = \mathbf{Q}_1 \mathbf{R}_1$ (QR-decomposition)

end

$\lambda_j = \|\mathbf{S}_1[:, j]\|_2$ for $j = 1, \dots, k$

$\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_k)$

for $\tau = 2 \rightarrow n$ **do**

$\mathbf{Q}_\tau = \mathbf{Q}_{\tau-1}$

$\tilde{\mathbf{X}}_\tau = \mathbf{M}_n^{\frac{1}{2}} \mathbf{X}_\tau$

for $i = 1 \rightarrow m$ **do**

$\mathbf{S}_\tau \leftarrow \frac{\tau-1}{\tau} \mathbf{Q}_{\tau-1} \Lambda_{\tau-1} \mathbf{Q}_{\tau-1}^\top \mathbf{Q}_\tau + \frac{1}{\tau} \frac{1}{B} \tilde{\mathbf{X}}_\tau^\top \tilde{\mathbf{X}}_\tau \mathbf{Q}_\tau$

$\mathbf{S}_\tau = \mathbf{Q}_\tau \mathbf{R}_\tau$ (QR-decomposition)

end

$\lambda_j = \|\mathbf{S}_\tau[:, j]\|_2$ for $j = 1, \dots, k$

$\Lambda_\tau = \text{diag}(\lambda_1, \dots, \lambda_k)$

end

Result: $\mathbf{Q}_\tau$

Then, the compression matrix $\mathbf{Q}_\tau$ obtained in Alg. 1. is used to compress the vibration data as shown in Eq. (20)

$$\tilde{\mathbf{X}}_c = \tilde{\mathbf{X}} \mathbf{Q}_\tau \quad (20)$$

, where $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1^\top, \tilde{\mathbf{X}}_2^\top, \dots, \tilde{\mathbf{X}}_n^\top]$ and $\tilde{\mathbf{X}}_c$ denotes the compressed data. As shown in Eq. (21), the vibration data can then be reconstructed from the compressed data.

$$\mathbf{X}_r = \mathbf{M}_n^{-\frac{1}{2}} \tilde{\mathbf{X}}_c \mathbf{Q}_\tau^\top \quad (21)$$

, where $\mathbf{X}_r$ denotes the reconstructed vibration data.

Benefit from the adaptive updating on eigenvectors, the computation complexity is dramatically reduced, and data compression efficiency is improved [20] in history PCA. The advantages of history PCA also reside in the improved rate of convergence [41]. Moreover, by introducing the normalized mass matrix into the data compression process, the approximation error can be further minimized as the model properties are better preserved. The experimental results introduced in the following two sections will validate the excellent performance.

V. NUMERICAL SIMULATION

In this section, data from a numerical arch bridge are used to validate the feasibility of the proposed data compression method. Procedures are introduced in detail, and the compressed vibration data (acceleration) were compared with the raw data. Its superiority over traditional history PCA-based method is also presented.

A. Numerical Arch Bridge

A 2D numerical arch bridge was built based on the material properties and geometric parameters shown in Table II. The vertical acceleration data of 28 points on the bridge were recorded with a sampling rate of 100Hz. An impact load was applied to the numerical arch bridge, and the free pulse responses of the 28 points on the bridge were obtained accordingly. The acceleration data were then arranged as a Hankel matrix, and the mode shapes are identified through ERA [27]. Afterward, the normalized mass matrix was obtained through solving the optimizing problem in Eq. (19) and used for the following data compression.

TABLE II
MATERIAL PROPERTIES AND GEOMETRIC PARAMETERS

Component	Arch rib	Main girder	Hanger
Material	C50	C40	1860 strand
Elastic modulus (Pa)	3.45×10^{10}	3.45×10^{10}	1.95×10^{11}
Poisson's ratio	0.2	0.2	0.3
Mass density (kg/m^3)	2549	2549×4	8005
Sectional area (m^2)	1.2	6	0.00567
Section moment of inertia (m^4)	0.2488	1.25	-
Section Height (m)	1.4	1	-
Section Width (m)	1.2	8	-

Typical daily wind forces were applied to the numerical arch bridge to simulate the ambient vibration of the arch bridge. Details of the numerical model and the wind velocity simulation methods have been introduced in [42], which are not the focus of this paper. The procedures for system identification and physics-enhanced PCA are summarized in Figure 4. Afterwards, the wind-induced responses at multiple locations were used to test the data compression algorithm. The raw acceleration data are premultiplied by $M_n^{-\frac{1}{2}}$ and then compressed by the physics-enhanced PCA.

B. Data Compression Results and Discussion

Data compression results on both raw and rescaled time series acceleration data of the 28 sensors are presented in this section for comparing the performance of physics-enhanced PCA and traditional history PCA. The comparison of raw and compressed acceleration collected by sensor 5 (on the main girder) and sensor 20 (on the arch) is shown in Figure 5. It illustrates visually that the compressed acceleration data approximate the raw data better when acceleration data are rescaled.

Signal to noise ratio (SNR) is also used to estimate the approximation error in representing the raw data with the two methods. And it is computed based on the following equation:

$$\text{SNR} = 20 \log_{10} \left(\frac{\|x\|_2}{\|x - \hat{x}\|_2} \right) \quad (22)$$

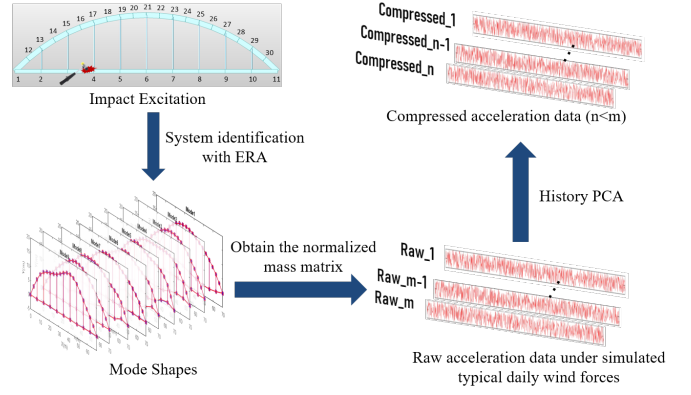


Fig. 4. Procedure of rescaling history PCA based data compression for the structure

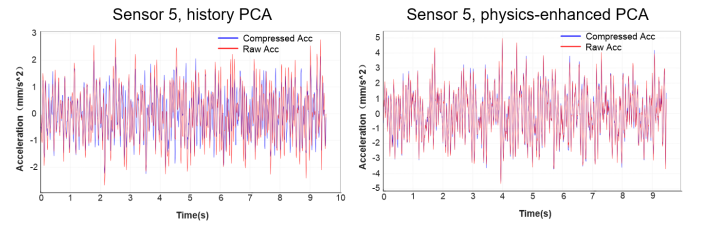


Fig. 5. Comparison of raw and compressed acceleration with and without rescaling

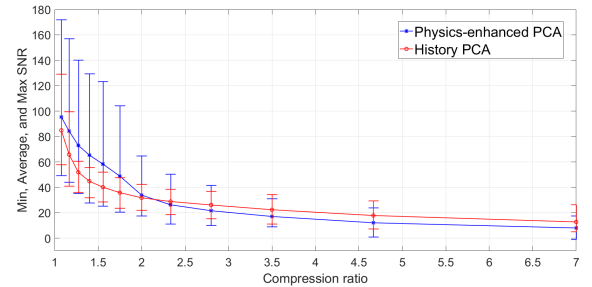


Fig. 6. Min, average, and max SNR between raw and compressed acceleration under different damage scenarios and compression ratio

The minimum, average, and maximum SNR for the 28 sensors were computed and presented in Figure 6. With the increase of compression ratio, the SNR for both methods decreases accordingly. When the compression ratio is less than 2, the SNR of physics-enhanced PCA exceeds history PCA both in average value and maximum value. For the minimum value of SNR, there is no significant difference between the two methods. When the compression ratio is larger than 2, the SNR values for both methods are pretty close. Because the larger SNR represents the less approximation error, it validates that the reconstructed acceleration data are closer to raw data when using physics-enhanced PCA.

To prove that the modal properties are preserved better in the proposed method. Frequencies and mode shapes identified from compressed acceleration data with the proposed method and history PCA are compared in Table III. Modal assurance criteria (MAC) is used to evaluate the error between mode

TABLE III
COMPARISON OF FREQUENCIES MODE SHAPES

Mode number	Frequencies (cycle/sec)			MAC	
	Raw acc	Compressed acc with HPCA	Compressed acc with PPCA	Compressed acc with HPCA	Compressed acc with PPCA
1	1.078	1.074 (0.37%)	1.078 (0.37%)	0.9995	0.9995(0.00%)
2	1.491	1.514 (1.54%)	1.514 (3.02%)	0.9390	0.9852(4.92%)
3	2.682	2.686 (0.15%)	2.686 (0.15%)	0.9806	0.9967(1.64%)
4	4.317	4.346 (0.67%)	4.346 (0.67%)	0.8698	0.9926(14.12%)
5	6.675	6.689 (0.21%)	6.689 (0.21%)	0.9997	0.9988(-0.09%)
6	8.455	8.496 (0.48%)	8.496 (0.09%)	0.9777	0.9890(1.16%)
7	9.381	9.375 (0.06%)	9.375 (0.06%)	0.9994	0.9955(-0.39%)
8	10.210	10.210 (0.00%)	10.210 (0.00%)	0.9995	0.9914(-0.81%)
9	12.340	12.000 (0.32%)	12.000 (0.32%)	0.9969	0.9914(-0.55%)

Note: PPCA denotes physics-enhanced PCA and HPCA denotes history PCA. Values in brackets of the 3rd and 4th columns are the relative error of frequencies. Values in brackets of the last column are the increase of MAC from HPCA to PPCA.

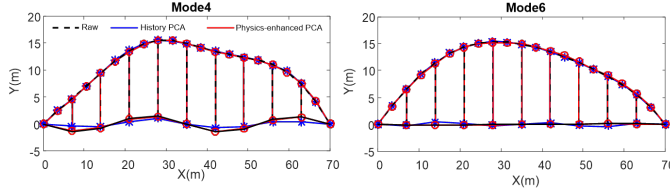


Fig. 7. Comparison of mode shapes of acceleration compressed with and without rescaling

shapes identified from raw and compressed acceleration. It is computed as below:

$$MAC = \frac{\varphi_i^T \varphi_j}{\varphi_j^T \varphi_j} \quad (23)$$

, where φ_j is the mode shape vector identified from raw acceleration and φ_i is the mode shape vector identified from compressed acceleration.

As Table III shows, frequencies are accurately preserved in compressed acceleration regardless of rescaling acceleration data or not. Results in MAC indicate that mode shapes are preserved more accurately when using physics-enhance PCA. The MAC increases up to 14.12% after using the proposed method compared with history PCA. Moreover, the fourth and sixth mode shapes are compared point-wise in Figure 7. Mode shapes identified in physics-enhanced PCA are closer to those identified from raw data.

VI. EXPERIMENT ON A REAL STRUCTURE

In this section, physics-enhanced PCA is tested with vibration data (acceleration) from a real-world structure. It runs in an edge computing based SHM system. The SHM system and data compression results are both introduced.

A. Experimental Setup

Acceleration responses from a two-story steel frame in [43] are used to validate the data compression performance of the proposed method. The cross-section of its column and beam are $50 \times 8.8 \text{ mm}^2$ and $50 \times 4.4 \text{ mm}^2$. The Young's modulus is $2.0 \times 10^5 \text{ MPa}$ and the mass density is $7.67 \times 10^3 \text{ kg/m}^3$. As shown in Figure 8, 14 accelerometers with 2048Hz sampling frequency are installed on the frame. Real-world structures are time-varying systems, and their structural modal properties, including frequencies and mode shapes, vary with the change

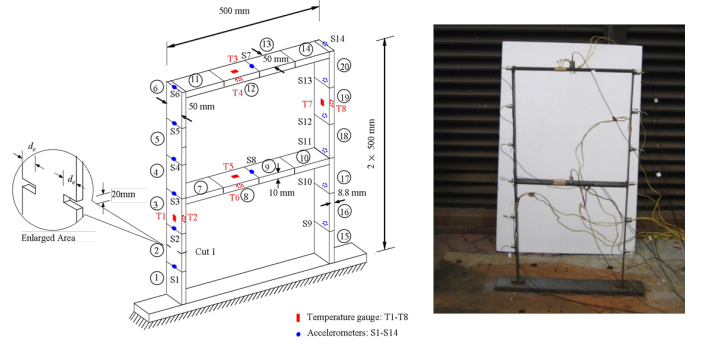


Fig. 8. Steel frame, sensor layout, and the location damage

of temperature and the degradation of the material. Also, structures are inevitably damaged during their long service period. To simulate those conditions, we performed different damage scenarios and collected corresponding vibration responses. More specifically, the steel frame was monitored from morning to afternoon, and the temperature varies from around 30 to around 60. After the acceleration responses under hammer impact were collected, a saw cut was introduced to simulate the damages, as shown in Figure 8. Three degrees of damage were simulated in this test, and the cut's depth (d_e) was set to 5, 10, and 15mm respectively. The length of each cut is 20mm. Because the saw cut is pretty short, the degradation in the mass of the element is negligible. However, the moment of inertia of the damage section was reduced by 20%, 40%, and 60% under the three degrees of damage, which would lead to a significant decrease in the stiffness of the structure.

The description of the eight types of damage scenario (DS) is summarized in Table IV. The first ten mode shapes obtained in damage scenario 1 are shown in Figure 9. The obtained modal matrix was then used to compute the normalized mass matrix.

TABLE IV
DAMAGE SCENARIOS FOR THE STEEL FRAME

Damage scenario	Damage condition	Data acquisition time
DS 1	Intact steel frame	7:55 AM
DS 2	Intact steel frame	12:10 PM
DS 3	Damaged steel frame, $d_e=5\text{mm}$	7:45 AM
DS 4	Damaged steel frame, $d_e=5\text{mm}$	12:05 PM
DS 5	Damaged steel frame, $d_e=10\text{mm}$	7:45 AM
DS 6	Damaged steel frame, $d_e=10\text{mm}$	12:10 PM
DS 7	Damaged steel frame, $d_e=15\text{mm}$	7:45 AM
DS 8	Damaged steel frame, $d_e=15\text{mm}$	12:05 PM

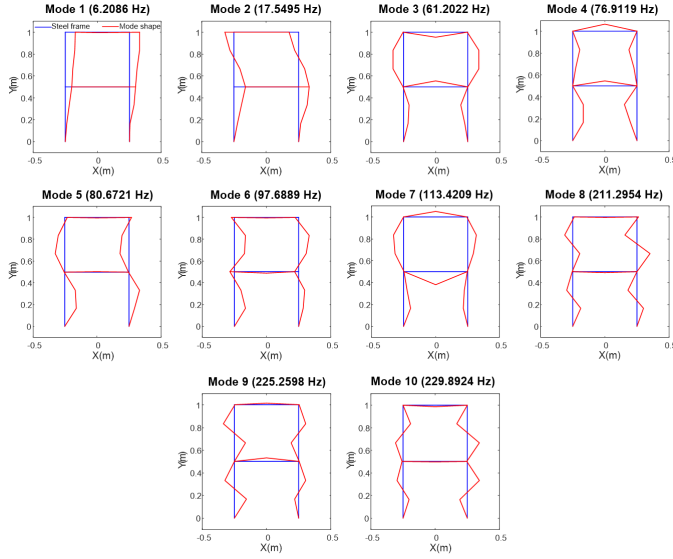


Fig. 9. Mode shapes for the intact steel frame

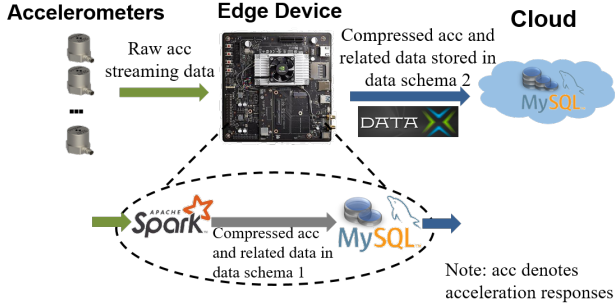


Fig. 10. Data transmission and processing procedure

B. Edge Computing Based Data Compression for a SHM System

The edge computing based SHM system consists of sensors, a Nvidia TX2, and a remote cloud platform. Equipped with 256-core NVIDIA Pascal GPU, Nvidia TX2 is one of the most advanced embedded computing devices for machine learning applications. Also, its Dual-Core NVIDIA Denver 2 64-Bit CPU and Quad-Core ARM Corte-A57 MPCore provide powerful computing resources for edge computing tasks. Physics-enhanced PCA algorithm is accomplished with Apache Spark [44], a streaming data processing engine widely used in cloud and edge computing. The algorithm is written in scala and Breeze library is used to support the matrix manipulation in this program. The acceleration data measured by each sensor were sent continuously to the port, which is listened by this algorithm. The data were then arranged into Dstream and processed by Spark. Afterward, the compressed data were sent to the cloud platform for backup. Authorized users who have remote access to the data can perform further damage identification or condition assessment. The flow chart for the data compression and transmission procedure is summarized in Figure 10.

As Table V shows, data stored in the edge device and cloud platform have different data schemas. And the definitions of

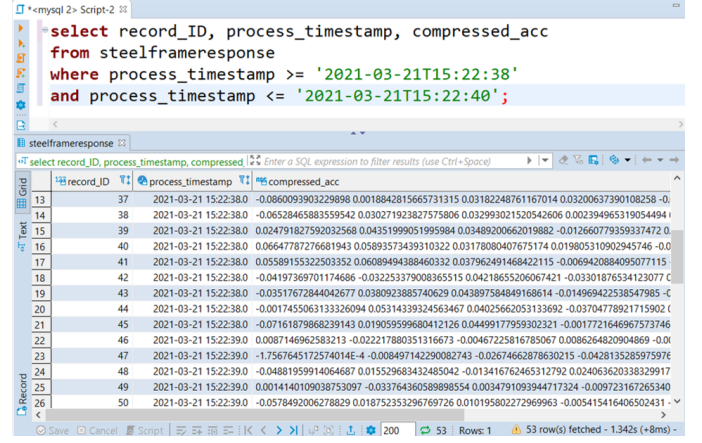


Fig. 11. Data retrieval and query result from database on the cloud

N, B, k, and block are the same with Algorithm 1. After the data compression, the compressed acceleration data and the raw acceleration data were both stored in the local Mysql database as the backup. Afterward, the compressed data $\tilde{\mathbf{X}}_c(t)$, timestamp, normalized mass matrix $\mathbf{M}_n$ and compression matrix $\mathbf{Q}_\tau$ were uploaded to the Mysql database in the remote cloud platform. To avoid frequent data transmission, the data transmission between edge and cloud is not in real-time. In this experiment, data were transmitted once a minute. Authorized users can retrieve data and query results as shown in Figure 11.

C. Analysis on the Data Compression Result

The normalized mass matrix computed from the acceleration responses of an intact steel frame is used to rescale the acceleration responses in eight damage scenarios. However, the mode shapes of the structure under different damage scenarios are different. To validate that the obtained normalized mass matrix can help preserve the modal properties under different damage scenarios, we compute Δ_n^2 with modal matrixes under the eight damage scenarios in this test. The computed Δ_n^2 for the eight scenarios are shown in Figure 12.

As shown in Figure 12, the normalized mass matrix obtained from intact steel frame remain orthogonal with respect to the modal matrix of steel frame under different temperature and damage level. So, it's unnecessary to update the normalized mass matrix frequently and the proposed approach can be easily performed in real project. Similar to the numerical model, both SNR and modal properties are used as metrics to evaluate the data compression performance. The average SNR is computed when using raw or compressed acceleration under the 8 damage scenarios in Table IV. As shown in Figure 13, the average SNR decreases with the increase of compression ratio. When compression ratio is less than 1.27, the SNR of physics-enhanced method is larger than history PCA. But the reverse applies when compression ratio lies between 1.27 and 2. When compression ration is larger than 2, the computed SNRs using history PCA and physics-enhanced PCA are very close. To sum up, physics-PCA and history PCA have similar

TABLE V
DATA SCHEMA FOR DATABASES IN THE EDGE DEVICE AND CLOUD PLATFORM

Data Schema 1		Data Schema 2		Description
Column_name	Data Type	Column_name	Data Type	
record_ID	INT	record_ID	INT	An int number to describe the number of record
process_timestamp	TIMESTAMP	process_timestamp	TIMESTAMP	Timestamp for each data compression operation
q_matrix	TEXT	q_matrix	TEXT	Each record consists of $N \times k$ elements in compression matrix $\mathbf{Q}_T$
compressed_acc	TEXT	compressed_acc	TEXT	Each record consists of $k \times B$ samples of compressed acceleration in one block.
rescaling_matrix	TINYTEXT	rescaling_matrix	TINYTEXT	Each record consists of the diagonal elements in the normalized mass matrix
sample_timestamp	TEXT	/	/	A string consists of the timestamp for each sample of raw acceleration
raw_acc	TEXT	/	/	Each record consists of $N \times B$ samples of raw acceleration in one block.

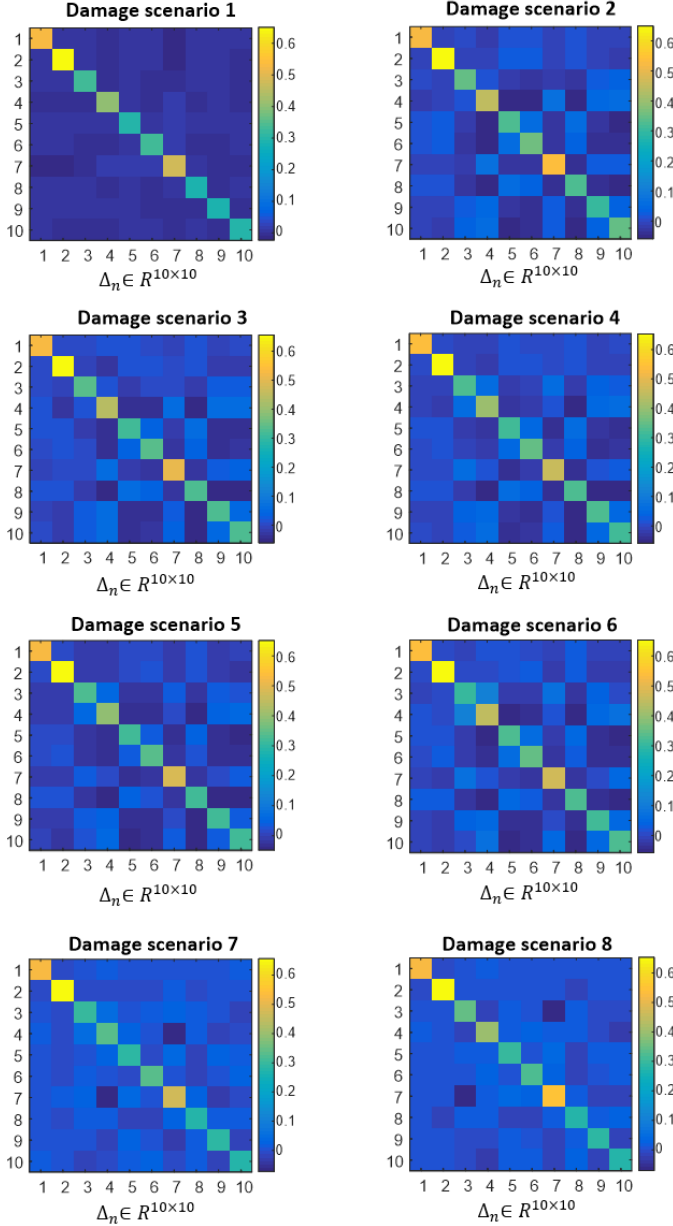


Fig. 12. Δ_n^2 computed with the identified mode shapes in the eight damage scenarios

reconstruction error in this experiment. Besides, the SNR for 8 damage scenarios are very close, which indicate the robustness of physics-PCA under different environmental and damage conditions.

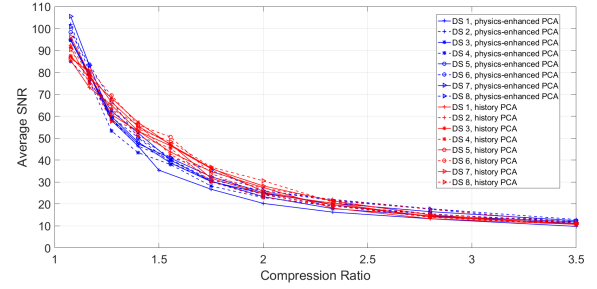


Fig. 13. Average SNR between raw and compressed acceleration under different damage scenarios and compression ratio

Frequencies and mode shapes identified from raw and compressed acceleration under 8 damage scenarios are compared in Table VI and Table VII. The values of CR in physics-enhanced PCA and history PCA are both set to 1.4 in this test. As listed in Table VI, the average values for relative error of frequencies are very close and all less than 0.2% in both methods, which shows the similar performance of the two methods in preserving frequencies. However, their performances in preserving mode shapes are quite different. Same as the last section, MAC is used to estimate the similarity between mode shapes computed from raw and compressed acceleration. As Table VII shows, the MAC values computed in physics-enhanced PCA are larger than history PCA for almost all damage scenarios and mode shapes with an average increment of 0.13% ~ 56.04%. The MAC value with the largest difference between the two methods is indicated in bold for each damage scenario. The mode shapes are compared point-wise in Figure 14 as well. Apparently, when physics-enhanced PCA is used to compress acceleration data, amplitudes of mode shapes extracted from the compressed data are much closer to those extracted from raw acceleration.

VII. CONCLUSION

In this paper, we have proposed a physics-enhanced PCA for compressing structural vibration data in edge devices. Considering the limited resources in edge devices, we innovatively integrate physical knowledge in data compression for compressing data while maintaining modal properties. Theoretical derivation and experiments results validate that the proposed method can preserve the mode shapes accurately while mitigating the bandwidth burden for edge devices. The monitoring of engineering structures in smart cities can be much more efficient and accurate. Compared with other data

TABLE VI
RELATIVE ERROR OF FREQUENCIES IDENTIFIED FROM RAW AND COMPRESSED ACCELERATION IN DIFFERENT DAMAGE SCENARIOS

Mode number	Relative error of frequencies															
	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA
mode1	0.05%	0.04%	0.01%	0.04%	0.14%	0.05%	0.13%	0.08%	0.00%	0.04%	0.05%	0.11%	0.03%	1.24%	0.12%	0.05%
mode2	0.04%	0.01%	0.02%	0.02%	0.03%	0.01%	0.03%	0.04%	0.01%	0.01%	0.01%	0.01%	0.23%	0.46%	0.02%	0.01%
mode3	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%
mode4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.03%	0.02%	0.04%	0.00%
mode5	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%
mode6	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%	0.01%	0.00%	0.05%	0.01%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%
mode7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.18%	0.24%	0.00%	0.01%	0.00%	0.22%	0.00%	0.01%	0.02%	0.05%
mode8	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.01%
mode9	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
mode10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.13%	0.00%
average	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.04%	0.04%	0.01%	0.01%	0.01%	0.04%	0.03%	0.17%	0.03%	0.01%

Note: The column of 'PPCA' denotes the result with physics-enhanced PCA. The column of 'HPCA' denotes the result with history PCA.

TABLE VII
MAC OF MODE SHAPES IDENTIFIED FROM RAW AND COMPRESSED ACCELERATION IN DIFFERENT DAMAGE SCENARIOS

Mode number	MAC of mode shapes															
	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8	
	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA	PPCA	HPCA
mode1	0.7303	0.7351	0.8465	0.5961	0.8854	0.5345	0.9408	0.9274	0.9479	0.9846	0.9264	0.9527	0.9861	0.7412	0.9844	0.9879
mode2	0.9690	0.3036	0.9973	0.9977	0.9954	0.9954	0.9975	0.9976	0.9997	0.9993	0.9997	0.9997	0.9966	0.9958	0.9998	0.9998
mode3	0.9998	0.9999	0.9994	0.9995	0.9998	0.9998	0.9997	0.9999	1.0000	0.9999	0.9997	0.9999	0.9999	0.9997	1.0000	1.0000
mode4	0.9994	0.9995	0.9995	0.9997	0.9995	0.9996	0.9527	0.1410	0.9996	0.9999	0.9997	0.9998	0.9992	0.9996	0.9976	0.9999
mode5	0.9999	0.9999	0.9996	0.9997	0.9998	0.9999	1.0000	1.0000	0.9999	1.0000	0.9999	0.9999	0.9987	0.9994	0.9997	1.0000
mode6	0.9941	0.9989	0.9054	0.9275	0.9969	0.9996	0.9944	0.9885	0.9981	0.9881	0.9903	0.9781	0.9895	0.9875	0.9748	0.9971
mode7	0.9142	0.9830	0.9798	0.9979	0.9930	0.9931	0.9049	0.9762	0.9796	0.3677	0.9966	0.9681	0.9676	0.9569	0.9556	0.2549
mode8	0.9992	0.9993	0.9994	0.9994	0.9996	0.9996	0.9985	0.9984	0.9980	0.9980	0.9968	0.9975	0.9990	0.9988	0.9982	0.9989
mode9	0.9994	0.9978	0.9999	1.0000	0.9999	0.9998	0.9996	0.9997	0.9998	0.9998	0.9996	0.9997	0.9998	0.9997	0.9995	1.0000
mode10	0.9981	0.9997	0.9908	0.9948	0.9995	0.9995	0.8954	0.9936	0.9779	0.9880	0.9976	0.9981	0.9992	0.9982	0.9339	0.9989
AG	21.10%		3.73%		6.53%		56.04%		16.27%		0.13%		3.45%		26.54%	

Note: PPCA denotes physics-enhanced PCA. HPCA denotes the history PCA. AG denotes average growth of MAC from HPCA to PPCA.

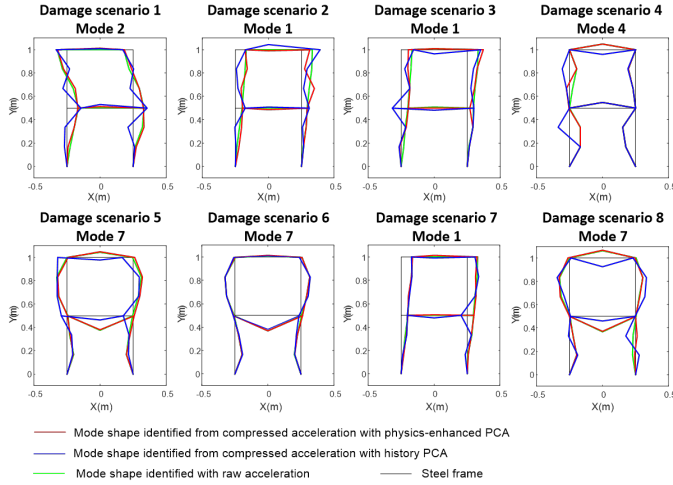


Fig. 14. Mode shapes identified from raw acceleration and compressed acceleration data

compression methods, the proposed method has the following advantages.

1. Its space and time complexity are far smaller than traditional PCA because data are processed incrementally in batches. Therefore, less power is consumed, and less computing resources are required for an edge device.

2. Guided by the normalized mass matrix M_n , the principal components extracted by PCA can be mapped into mode shapes of the structure. Supported by the theoretic proof and experiment results, the physics-enhance PCA can preserve

substantial physical quantities more accurately than traditional history PCA.

3. The normalized mass matrix M_n can be considered as time-invariant within a small period under normal operation. There is no need to update it frequently, which is convenient for real-word applications.

In addition, this paper presents an impressive practice in integrating physical knowledge with data analysis, which can be borrowed in many types of scientific analysis. Nevertheless, some limitations still exist in this method. It only works on the ambient vibration of a linear time-invariant system at present. Further work should be done for the adaptation to nonlinear systems or time-varying systems. Besides, its applicability in dealing with forced vibration data should be analyzed in the future.

ACKNOWLEDGEMENT

This work described in this paper was supported by grants from the GDSTC Key Technologies RD Programme (2019B111106001), and from the Research Grants Council of Hong Kong through the Theme-based Research Scheme (T22-502/18-R). The findings and opinions expressed in this paper are from the authors alone and are not necessarily the views of the sponsors.

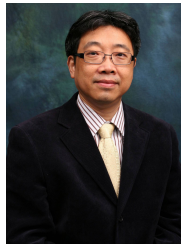
REFERENCES

- [1] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection," *Journal of Structural Engineering*, vol. 146, no. 5, p. 04020073, 2020.

- [2] S. Jeong, R. Hou, J. P. Lynch, H. Sohn, and K. H. Law, "A scalable cloud-based cyberinfrastructure platform for bridge monitoring," *Structure and Infrastructure Engineering*, vol. 15, no. 1, pp. 82–102, 2019.
- [3] L. Yu and J.-C. Lin, "Cloud computing-based time series analysis for structural damage detection," *Journal of Engineering Mechanics*, vol. 143, no. 1, p. C4015002, 2017.
- [4] D. Xu, Q. Li, and H. Zhu, "Energy-saving computation offloading by joint data compression and resource allocation for mobile-edge computing," *IEEE Communications Letters*, vol. 23, no. 4, pp. 704–707, 2019.
- [5] J. Azar, A. Makhoul, M. Barhamgi, and R. Couturier, "An energy efficient iot data compression approach for edge machine learning," *Future Generation Computer Systems*, vol. 96, pp. 168–175, 2019.
- [6] C. Chen, Y. Ding, Z. Wang, J. Zhao, B. Guo, and D. Zhang, "Vtracer: When online vehicle trajectory compression meets mobile edge computing," *IEEE Systems Journal*, vol. 14, no. 2, pp. 1635–1646, 2019.
- [7] S. Lu, Q. Xia, X. Tang, X. Zhang, Y. Lu, and J. She, "A reliable data compression scheme in sensor-cloud systems based on edge computing," *IEEE Access*, vol. 9, pp. 49 007–49 015, 2021.
- [8] T. Bicer, J. Yin, D. Chiu, G. Agrawal, and K. Schuchardt, "Integrating online compression to accelerate large-scale data analytics applications," in *2013 IEEE 27th International Symposium on Parallel and Distributed Processing*. IEEE, 2013, pp. 1205–1216.
- [9] A. H. Baker, H. Xu, J. M. Dennis, M. N. Levy, D. Nychka, S. A. Mickelson, J. Edwards, M. Vertenstein, and A. Wegener, "A methodology for evaluating the impact of data compression on climate simulation data," in *Proceedings of the 23rd international symposium on High-performance parallel and distributed computing*, 2014, pp. 203–214.
- [10] S. X. Wu, H.-T. Wai, L. Li, and A. Scaglione, "A review of distributed algorithms for principal component analysis," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1321–1340, 2018.
- [11] K. R. Rao and P. C. Yip, *The transform and data compression handbook*. CRC press, 2018.
- [12] F. Ni, J. Zhang, and M. N. Noori, "Deep learning for data anomaly detection and data compression of a long-span suspension bridge," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 685–700, 2020.
- [13] D. Laney, S. Langer, C. Weber, P. Lindstrom, and A. Wegener, "Assessing the effects of data compression in simulations using physically motivated metrics," in *SC'13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. IEEE, 2013, pp. 1–12.
- [14] N. Sasaki, K. Sato, T. Endo, and S. Matsuoka, "Exploration of lossy compression for application-level checkpoint/restart," in *2015 IEEE International Parallel and Distributed Processing Symposium*. IEEE, 2015, pp. 914–922.
- [15] Z. Lu, D. Y. Kim, and W. A. Pearlman, "Wavelet compression of ecg signals by the set partitioning in hierarchical trees algorithm," *IEEE transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 849–856, 2000.
- [16] N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, "A wireless sensor network for structural monitoring," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, 2004, pp. 13–24.
- [17] S. Liu and L. Cheng, "Efficient data compression in wireless sensor networks for civil infrastructure health monitoring," in *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, vol. 3. IEEE, 2006, pp. 823–829.
- [18] J. Ning, J. Wang, W. Gao, and C. Liu, "A wavelet-based data compression technique for smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 212–218, 2010.
- [19] Y. Bao, J. L. Beck, and H. Li, "Compressive sampling for accelerometer signals in structural health monitoring," *Structural Health Monitoring*, vol. 10, no. 3, pp. 235–246, 2011.
- [20] A. Burrello, A. Marchioni, D. Brunelli, S. Benatti, M. Mangia, and L. Benini, "Embedded streaming principal components analysis for network load reduction in structural health monitoring," *IEEE Internet of Things Journal*, 2020.
- [21] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [22] T. Elgamal and M. Hefeeda, "Analysis of pca algorithms in distributed environments," *arXiv preprint arXiv:1503.05214*, 2015.
- [23] G. A. Korn, *Interactive Dynamic-System Simulation*. Crc Press, 1998.
- [24] M. De Angelis, H. Lus, R. Betti, and R. W. Longman, "Extracting physical parameters of mechanical models from identified state-space representations," *J. Appl. Mech.*, vol. 69, no. 5, pp. 617–625, 2002.
- [25] D. M. Luchtenburg and C. W. Rowley, "Model reduction using snapshot-based realizations," *Bulletin of the American Physical Society*, vol. 56, no. 18, 2011.
- [26] A. Quarteroni, G. Rozza *et al.*, *Reduced order methods for modeling and computational reduction*. Springer, 2014, vol. 9.
- [27] J.-N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *Journal of guidance, control, and dynamics*, vol. 8, no. 5, pp. 620–627, 1985.
- [28] C. W. Rowley, "Model reduction for fluids, using balanced proper orthogonal decomposition," *International Journal of Bifurcation and Chaos*, vol. 15, no. 03, pp. 997–1013, 2005.
- [29] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.
- [30] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.
- [31] S. Han and B. Feeny, "Application of proper orthogonal decomposition to structural vibration analysis," *Mechanical Systems and Signal Processing*, vol. 17, no. 5, pp. 989–1001, 2003.
- [32] C. Wang, W. Guan, J. Wang, B. Zhong, X. Lai, Y. Chen, and L. Xiang, "Adaptive operational modal identification for slow linear time-varying structures based on frozen-in coefficient method and limited memory recursive principal component analysis," *Mechanical Systems and Signal Processing*, vol. 100, pp. 899–925, 2018.
- [33] R. Clough and J. Penzien, "Dynamics of structures," Tech. Rep., 1975.
- [34] L. Yu and H. Song, "Scaling mode shapes in output-only structure by a mass-change-based method," *Shock and Vibration*, vol. 2017, 2017.
- [35] Z. Ge and Z. Song, "Process monitoring based on independent component analysis- principal component analysis (ica- pca) and similarity factors," *Industrial & Engineering Chemistry Research*, vol. 46, no. 7, pp. 2054–2063, 2007.
- [36] H. Lus, "Control theory based system identification." 2002.
- [37] J. Cara, "Computing the modal mass from the state space model in combined experimental-operational modal analysis," *Journal of Sound and Vibration*, vol. 370, pp. 94–110, 2016.
- [38] G. Acunzo, N. Fiorini, F. Mori, and D. Spina, "Modal mass estimation from ambient vibrations measurement: A method for civil buildings," *Mechanical Systems and Signal Processing*, vol. 98, pp. 580–593, 2018.
- [39] A. Sestieri and S. Ibrahim, "Analysis of errors and approximations in the use of modal co-ordinates," *Journal of sound and vibration*, vol. 177, no. 2, pp. 145–157, 1994.
- [40] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *Journal of mathematical analysis and applications*, vol. 106, no. 1, pp. 69–84, 1985.
- [41] P. Yang, C.-J. Hsieh, and J.-L. Wang, "History pca: A new algorithm for streaming pca," *arXiv preprint arXiv:1802.05447*, 2018.
- [42] Y. Duan, Q. Chen, H. Zhang, C. B. Yun, S. Wu, and Q. Zhu, "Cnn-based damage identification method of tied-arch bridge using spatial-spectral information," *Smart Structures and Systems*, vol. 23, no. 5, pp. 507–520, 2019.
- [43] Y. Bao, Y. Xia, H. Li, Y.-L. Xu, and P. Zhang, "Data fusion-based structural damage detection under varying temperature conditions," *International journal of structural stability and dynamics*, vol. 12, no. 06, p. 1250052, 2012.
- [44] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin *et al.*, "Apache spark: a unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.



Qianyi Chen is currently a Ph.D. candidate in the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China. Before that, he holds a M.S. (2019) from Zhejiang University and a B.S. (2016) from Huazhong University of Science and Technology. His research interests include physics-enhanced machine learning, edge computing, and structural health monitoring.



Jiannong Cao received his M.Sc. and Ph.D. degrees in computer science from Washington State University. He is currently a Chair Professor with the Department of Computing at The Hong Kong Polytechnic University (PolyU). He is also the dean of Graduate School, the director of Research Institute of Artificial Intelligent of Things, the director of the Internet and Mobile Computing Lab and the vice director of the University's Research Facility in Big Data Analytics in PolyU. His research interests include distributed systems and blockchain, wireless

sensing and networking, big data and machine learning, and mobile cloud and edge computing. He has co-authored 5 books, co-edited 9 books, and published over 500 papers in major international journals and conference proceedings. He is a member of Academia Europaea, fellow of IEEE, and an ACM distinguished member.



Yong Xia holds a B.S. (1992) and M.S. (1997) from Huazhong University of Science and Technology and a Ph.D. (2002) from the Nanyang Technological University. He is the Associate Head and Professor at the Department of Civil and Environmental Engineering, the Hong Kong Polytechnic University. His primary research interests include structural health monitoring and structural dynamics. He has co-authored over 130 refereed journal papers, one research book, five design standards, and four conference proceedings. He is now the Fellow of the

Hong Kong Institution of Engineers and Co-Editor-in-Chief of *Advances in Structural Engineering*.