The application of unsupervised learning to a dataset of AC susceptibility measurements of HTS

M. Kowalik[1,2] (marcin@aisolutions.pl), R. Zalecki[2], M. Giebułtowski[2], W. Tokarz[2], J. Niewolski[2]

[1]- Department of Physics and Medical Engineering, Rzeszów University of Technology, al. Powstańców Warszawy 6, 35-959 Rzeszów, Poland

[2]– Department of Solid State Physics, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland

INTRODUCTION

High-temperature superconductors (HTS) are materials which exhibit properties like zero electrical resistance and expulsion of an external magnetic field from the interior of superconductor. These unique properties are used in the state of the art applications for the medicine (MRI/NMR machines), science (particle accelerators), transportation (Maglev - magnetic levitation train) and electrical industry (electric power transmission, fault current limiters). The greatest drawback of superconductors is that they only function in low temperatures. Scientist are still pursuing the discovery of room-temperature superconductor. The phenomenon of high-temperature superconductivity is still not fully understood.

Our work aims to provide insights if clustering technique applied to the dataset consisting of the measurements of High-Temperature Superconductors using the AC susceptibility method, will allow recovering known relationships (features) between different types of high-temperature superconductors and their superconducting properties, which depends on sample preparation conditions like sintering and annealing temperatures, etc.

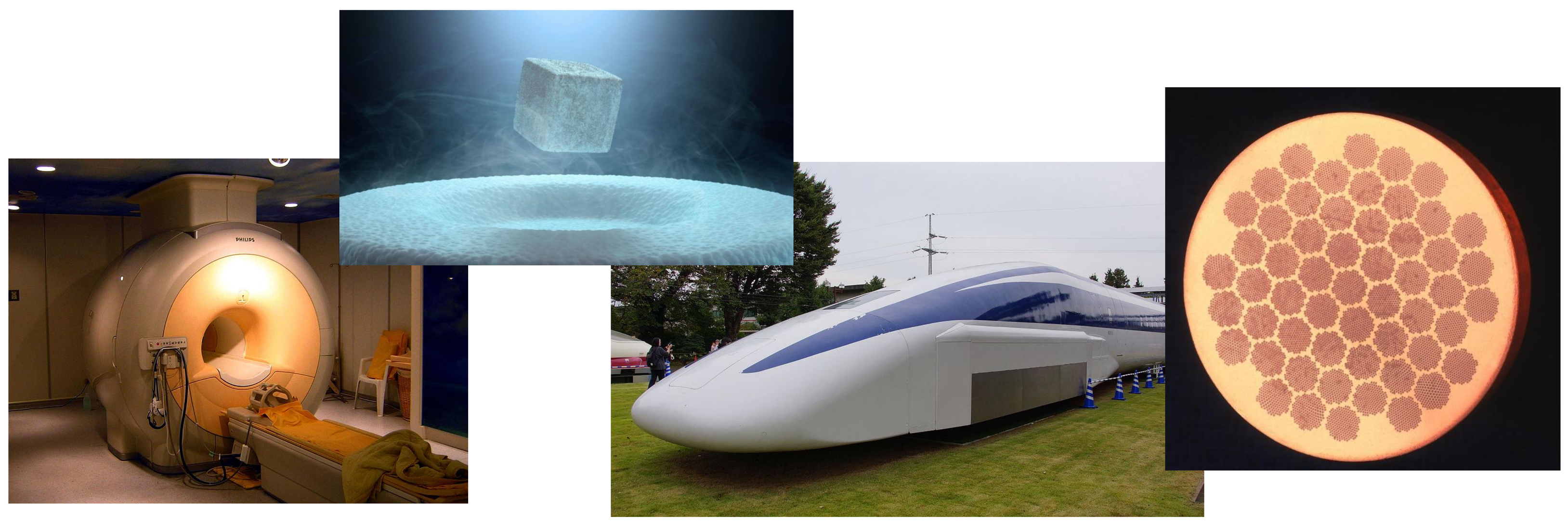


***Fig 1.** Modern applications of HTS*

BAG OF WORDS APPROACH

Measuring the temperature dependence of the complex AC susceptibility χ is the most common procedure for characterizing the properties like critical temperature T_c, critical current density j_c and granularity (microstructure) of HTS samples.

ASSUMPTION 1: The single AC susceptibility vs temperature measurement can be treated as a sequence of several hundred data points (**sentence**), therefore it can be represented as a collection of smaller subsequences (**words**).

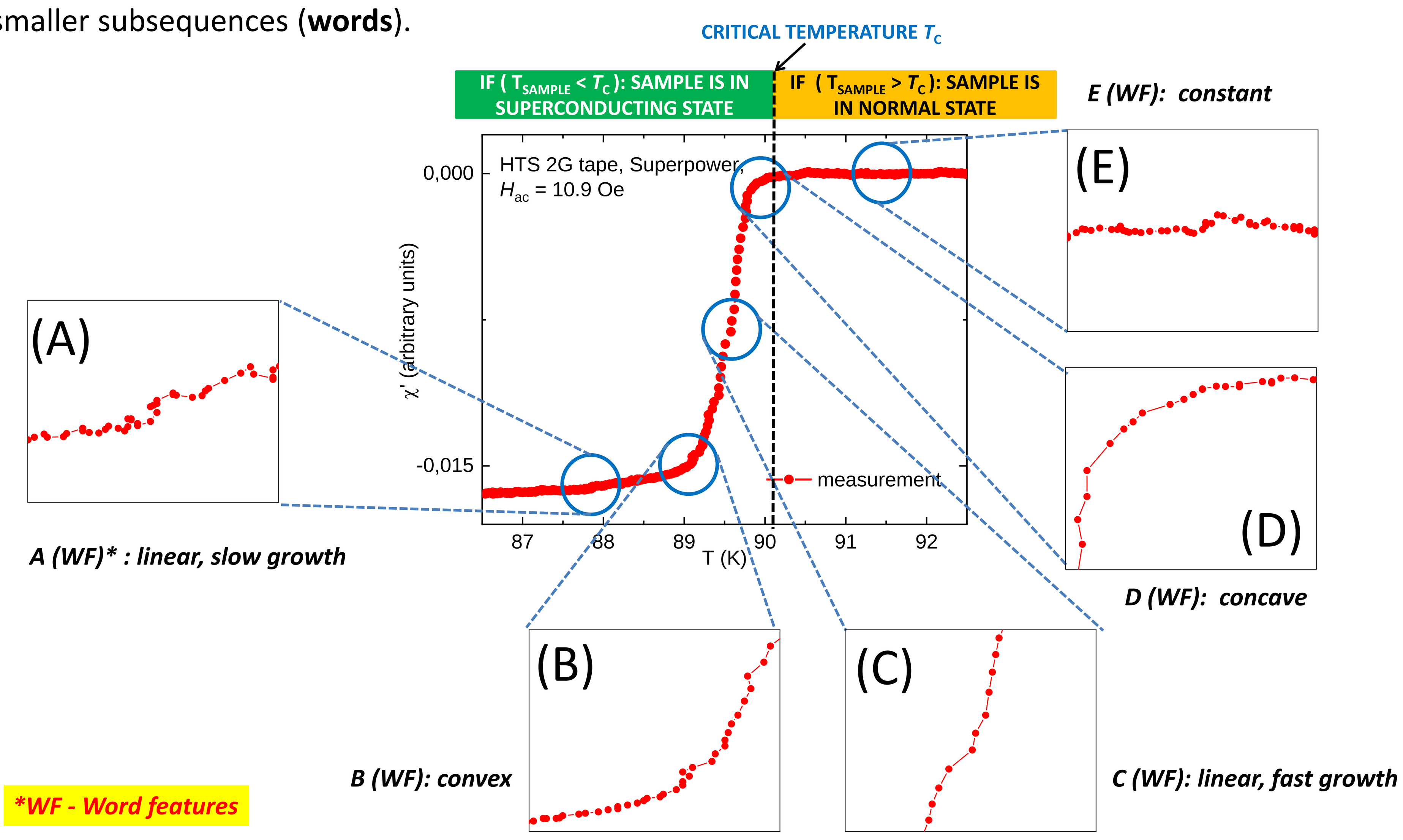


***Fig 2.** Bag of words approach in the analysis of AC measurement. Small figures marked as A, B, C, D and E represent the words of a sequence shown on big, middle plot.*

ASSUMPTION 2: The shape of AC susceptibility vs temperature measurement depends on the superconducting and microstru-cture features of measured HTS sample. Therefore constructing a histogram of word occurrences should allow for comparison of features of χ' mea-surements of different HTS samples.

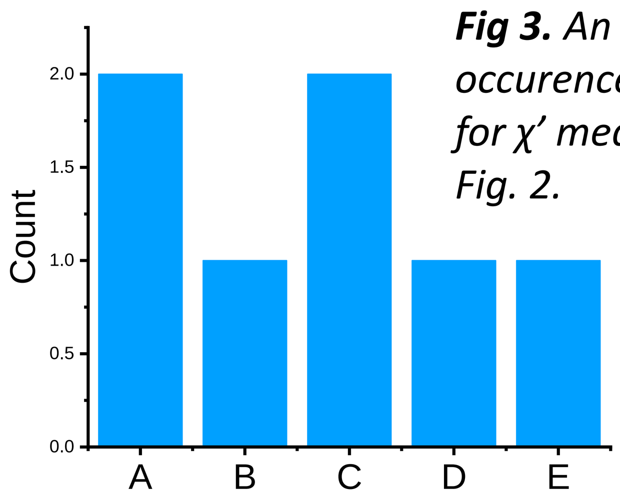


***Fig 3.** An example of word occurences histogram created for χ' measurement shown in Fig. 2.*

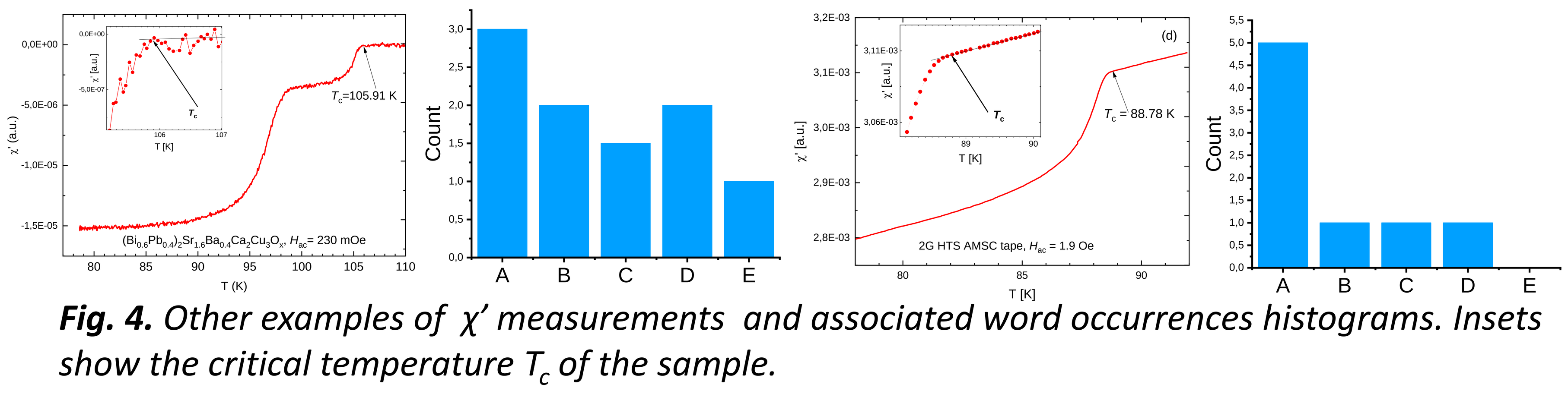
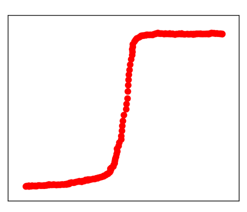
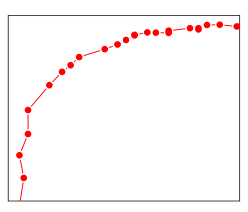


***Fig. 4.** Other examples of χ' measurements and associated word occurrences histograms. Insets show the critical temperature T_c of the sample.*

WHAT IS a SENTENCE and a WORD



Sentence: χ' measurement, con-sisting of on average about 400 data points. Each data point is 2D vector consisting current values of [sample temperature, value of χ']. The first coordinate (temperature) is ignored and the assumption is made that the distance between succeeding data points is always the same.



Word: vector of length of about several dozen succeeding values of χ'. In our case, the single word consists of 45 values. The words were created from the X' measurement by the sliding window method.

BAG OF WORDS IMPLEMENTATION TO THE DATASET

The dataset consists of about 1000 measurements of AC susceptibility of HTS samples, which results in about 300k words.

Calculation setps:

1. **Normalization of each word to values between 0 and 1.**
2. **Reduction of dimensionality of words (45 -> 9) using a CNN autoencoder**
3. **Clustering of encoded sequences by k-means algorithm to extract unique words**
4. **Visualisation of the unique words using t-SNE**
5. **Transformation of all of the χ' (T) measurements into 6D representation (FINAL RESULT)**

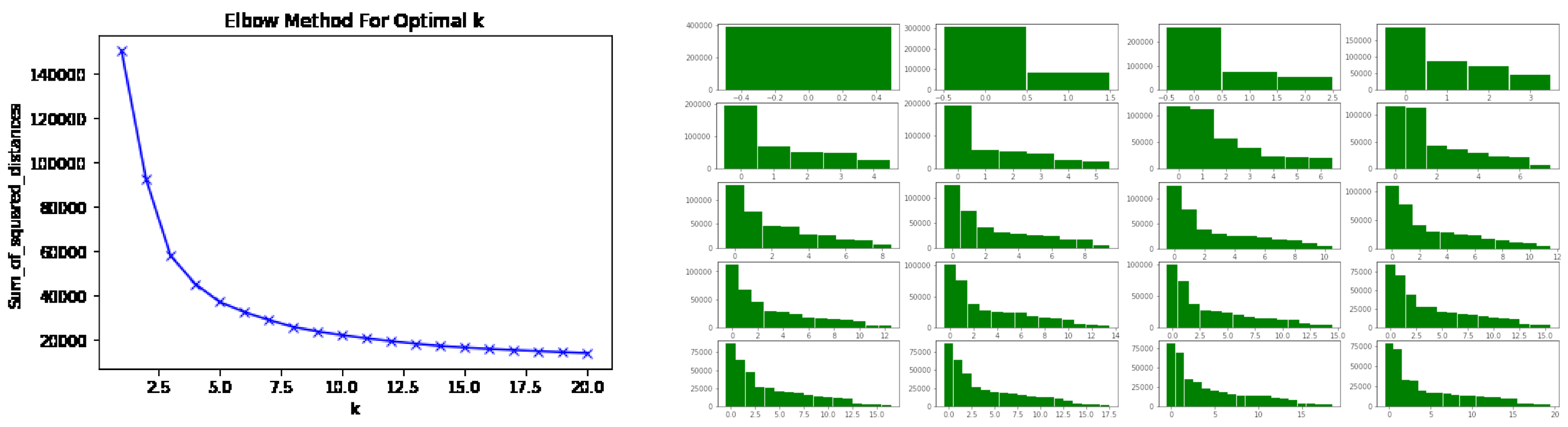


***Fig 5.** Estimation of the optimal cluster number for k-means algorithm. Left graph – The elbow method for an optimal K estimation, right graph - number of clusters and resulting clusters sizes.*

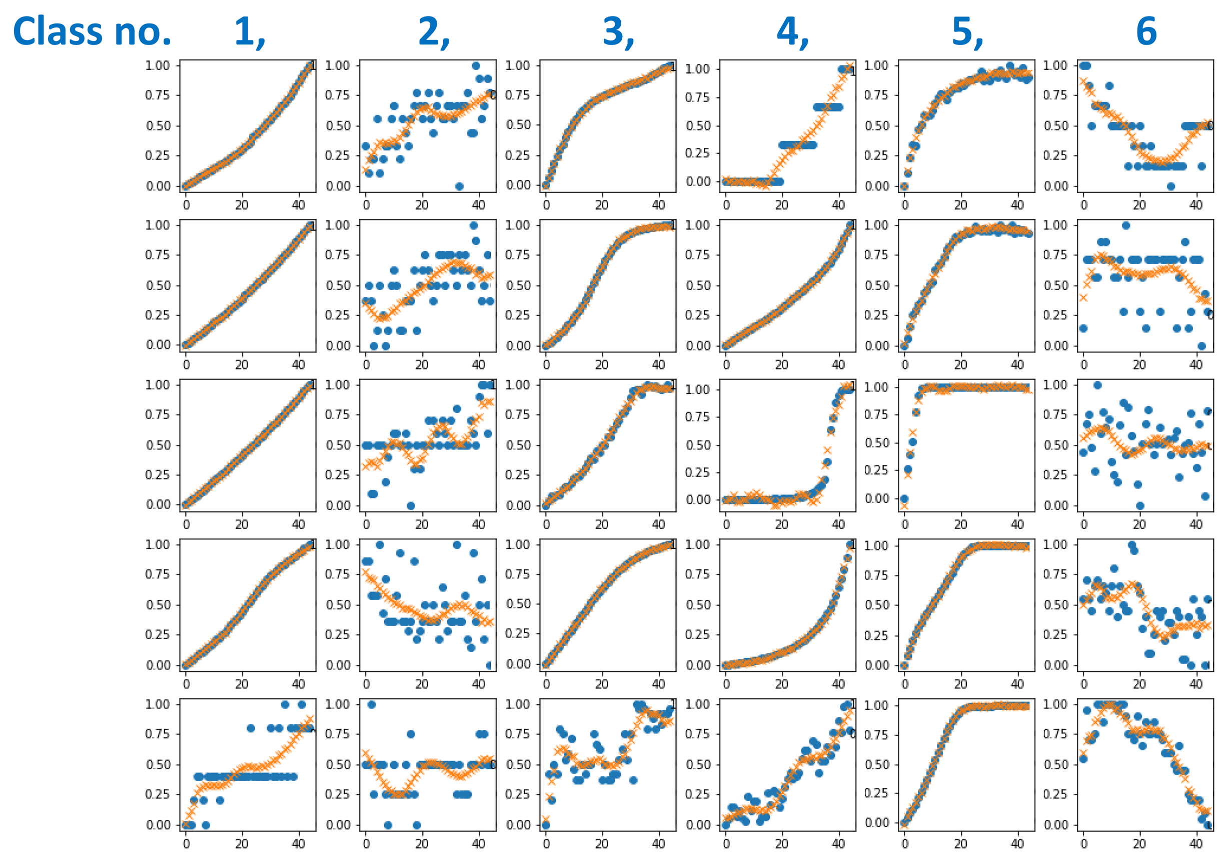


***Fig 6.** Result of clustering for randomly selected words into 6 classes. Blue dots represent the X' values included in a single row. Orange crosses show a word reconstructed by NN autoencoder.*

t-SNE visualisation of the words set grouped into 6 classes by k-means algorithm

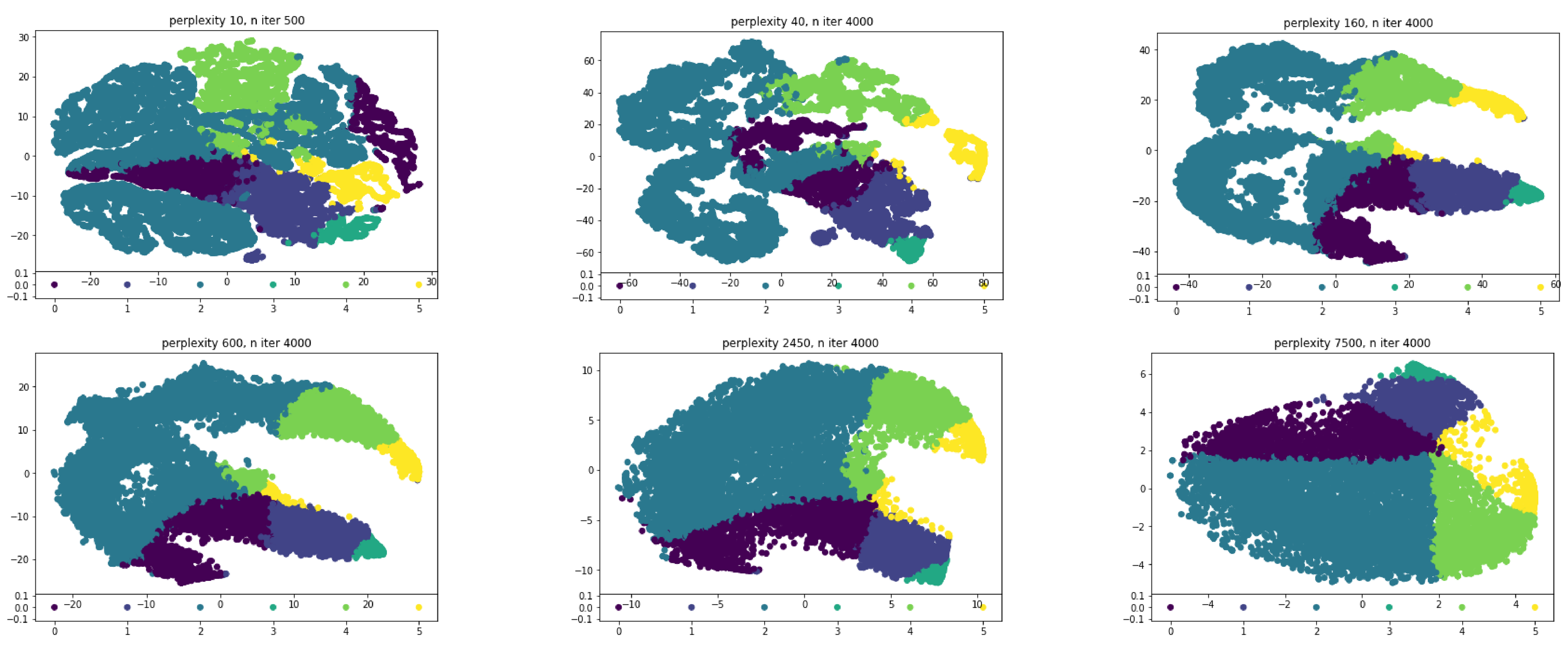


***Fig 7.** t-SNE visualisation of 20k random words for a number of clusters set to 6*

RESULTS

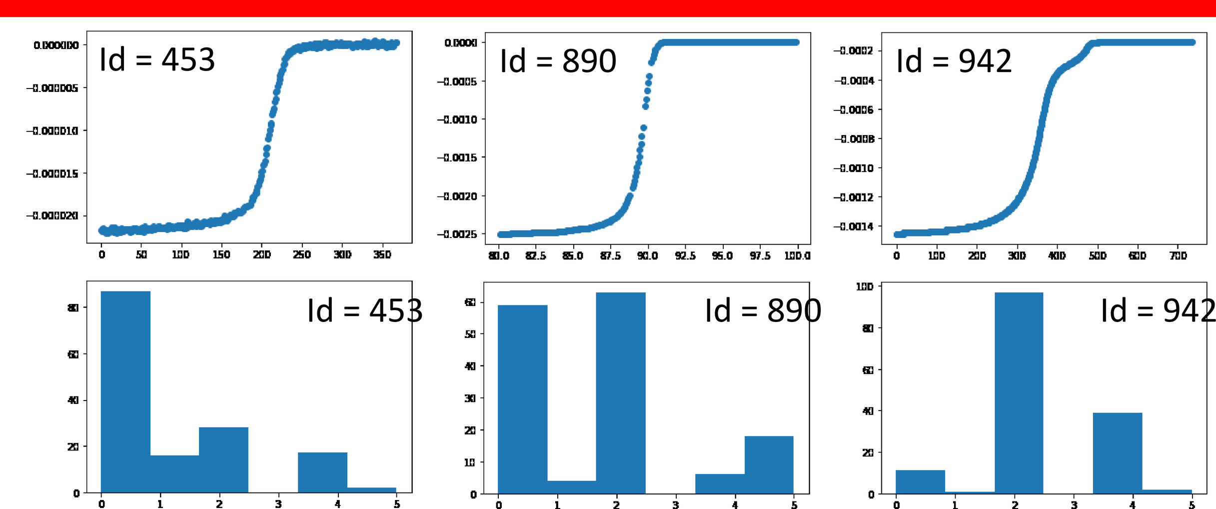


***Fig 8.** Randomly selected χ' measurements and associated word occurrences histograms.*

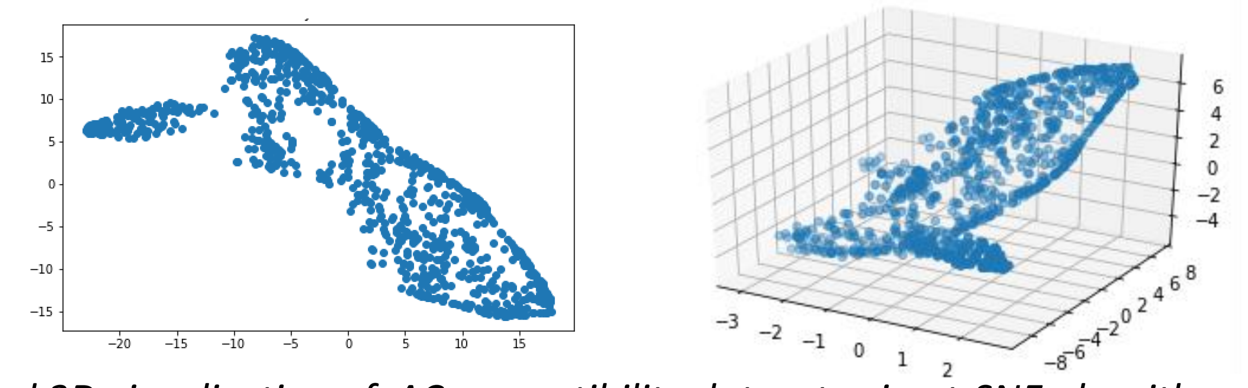


***Fig 9.** 2D and 3D visualisation of AC susceptibility dataset using t-SNE algorithm. A single point represents a single χ' (T) measurement. Axes represents t-SNE features.*

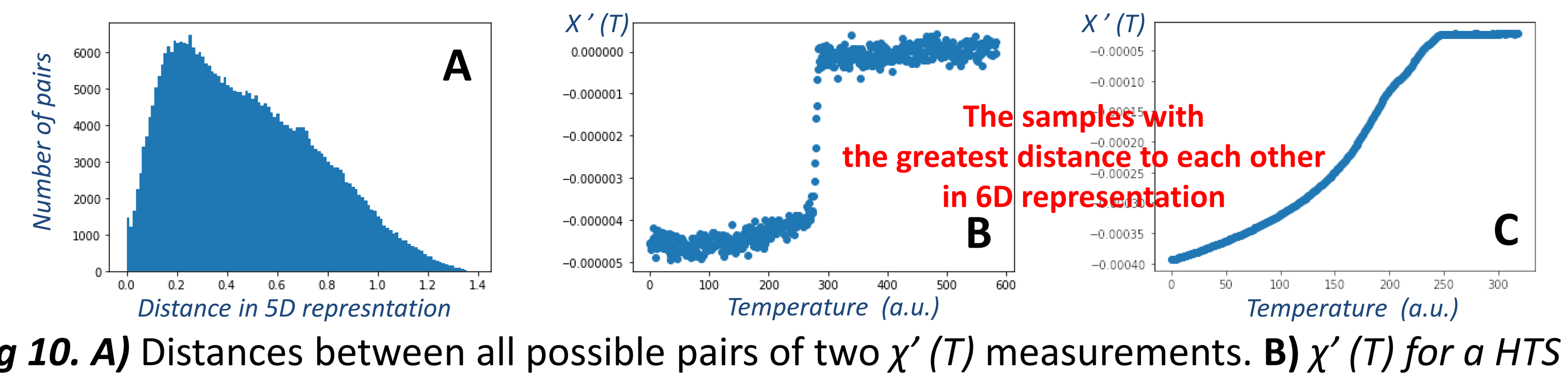


***Fig 10. A)** Distances between all possible pairs of two χ' (T) measurements. **B)** χ' (T) for a HTS wire (high value of critical current density, sharp S.C. transition) **C)** χ' (T) for grinded and pressed into a pellet polycrystalline REBCO-123 (electric current can not flow, wide SC transition)*

Acknowledgement: This work is supported by The Polish National Science Center under grant No. DEC-2018/02/X/ST3/01741