Preprints.org

Article

Not peer-reviewed version

Unlocking the Hidden Potential of the Data-Driven Nanocarbon Genome: Unleashing Novel Neural Pathways, Elucidating Growth-Property Relationships, and Enabling Inverse Design

Alexander Lukin * and Oğuz Gülseren

Posted Date: 20 February 2024

doi: 10.20944/preprints202402.1117.v1

Keywords: low-dimensional nanocarbons; nano-informatics; data-driven nanocarbon genome approach; multifactorial predictive models; universal descriptors; growth-property relationships; data-driven neural pathways; nano-interfaces; collective atomic vibrations; phonon engineering; data-driven inverse design, cyber–physical systems

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Unlocking the Hidden Potential of the Data-Driven Nanocarbon Genome: Unleashing Novel Neural Pathways, Elucidating Growth-Property Relationships, and Enabling Inverse Design

Alexander Lukin [1,*] and Oğuz Gülseren [2]

[1] Western-Caucasus Research Center, Tuapse 352815, Russian Federation; lukin@wcrc.ru
[2] Department of Physics, Bilkent University, Ankara 06800, Turkey; gulseren@fen.bilkent.edu.tr
* Correspondence: lukin@wcrc.ru; Tel.: +7-918-3080916

Abstract: The swift progress in low-dimensional nanocarbons, specifically in the realm of the authentic one-dimensional carbon chain featuring sp^1 hybridization, has unveiled exciting prospects for integrating them into cutting-edge technologies across diverse industries. To fully harness their potential, precise control over the growth process is necessary to obtain desired nanostructure and functionality. However, optimizing properties through traditional approaches has been challenging due to complex interactions. To address this, we implement a focused data-driven strategy for nanocarbon inverse design by leveraging a state-of-the-art data-driven nanocarbon genome approach (NCGA). By uncovering relationships between growth parameters and resultant traits, this serves as an expedient catalyst in engineering nanostructures with tailored attributes. We introduce an extensive array of technological approaches aimed at precisely controlling the growth process. These methods encompass stimulating and precisely adjusting synergistic effects, coordinating atomic vibrations at a collective level through the implementation of multilayer nano-interfaces, harnessing the potential of active screen plasma surface engineering, utilizing nano-patterning and allotropic phase transformations, incorporating heteroatom doping, and effectively directing self-assembly. These aim to unlock nanomaterials' latent potential and reveal novel neural pathways within the data-driven NCGA, enhancing its predictive capabilities. Specifically, triggering self-organization during growth can potentially unlock previously unexplored neural pathways. The data-driven NCGA offers a paradigm shift, accelerating discovery and design of nanocarbons with optimized, application-specific properties.

Keywords: low-dimensional nanocarbons; nano-informatics; data-driven nanocarbon genome approach; multifactorial predictive models; universal descriptors; growth-property relationships; data-driven neural pathways; nano-interfaces; collective atomic vibrations; phonon engineering; data-driven inverse design; cyber–physical systems

1. Introduction

The field of nanomaterials science has witnessed remarkable progress in recent times, primarily fueled by the significant contributions made by low-dimensional carbon nanomaterials, [1]. These carbon allotropes possess unique nanostructures that confer distinctive physicochemical properties and functional capabilities, [2]. With their versatility, low-dimensional carbon nanomaterials can be seamlessly integrated into hybrid systems and multifunctional composites, resulting in the creation of materials that exhibit exceptional qualities. By tapping into the power of low-dimensional nanocarbon allotropes, the functional properties of various nano-materials in advanced technologies can be maximized, unlocking their full potential.

Among one-dimensional carbon allotropes, carbyne represents the ultimate linear carbon chain and coveted holy grail, [3–5]. Comprised of sp-hybridized carbon atoms linked in linear chains with alternating double and triple bonds, carbyne is the thinnest possible form of carbon. With exceptional

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

tensile strength and stiffness exceeding diamond and graphene, carbyne is renowned as one of the strongest materials known. Beyond its superlative mechanical properties, carbyne also exhibits unique electrical characteristics. Under applied tension, carbyne can transform remarkably from a conducting cumulene form to insulating polyyne. This ability to toggle conductivity makes carbyne an intriguing functional material. Additionally, the optical absorption spectrum changes significantly with applied strain, allowing precise control over sensitivity to light.

By conducting research on carbyne and its formation mechanism, it becomes possible to gain a more profound comprehension of carbon's behavior in extreme situations. Furthermore, this exploration unveils potential applications for these exceptional structures.

The instability and high reactivity of carbyne - a unique form of carbon - pose challenges for the growth of large carbyne crystals. Ensuring strong stability of this nanomaterial becomes paramount for the practical applications of carbyne-like nanostructures.

The synthesis of carbyne and related nanostructures continues to be a noteworthy obstacle, primarily because our understanding of the complex relationships between growth mechanisms and resulting properties remains limited, [6].

In recent times, a new technology has emerged to counteract the excessive reactivity of carbon chains, [7]. This breakthrough has given rise to a distinctive form of carbon called two-dimensionally ordered linear chain carbon (2D LCC). The development of this technology allowed for the incorporation of aligned carbon chains into an amorphous carbon matrix using ion-assisted pulsed plasma deposition. By implementing this method, the growth of carbyne chains is effectively facilitated within a secure environment. In addition, the transformation of carbyne into alternative carbon phases like graphite or diamond is limited due to the repulsion between carbon chains when they come into proximity as a result of van der Waals forces. The spatial configuration of the 2D LCC bears a resemblance to a distributed hexagonal array, where parallel one-dimensional carbon chains are arranged in a two-dimensional manner. These chains are connected by van der Waals forces and are oriented at a right angle to the substrate's surface, (Figure 1).

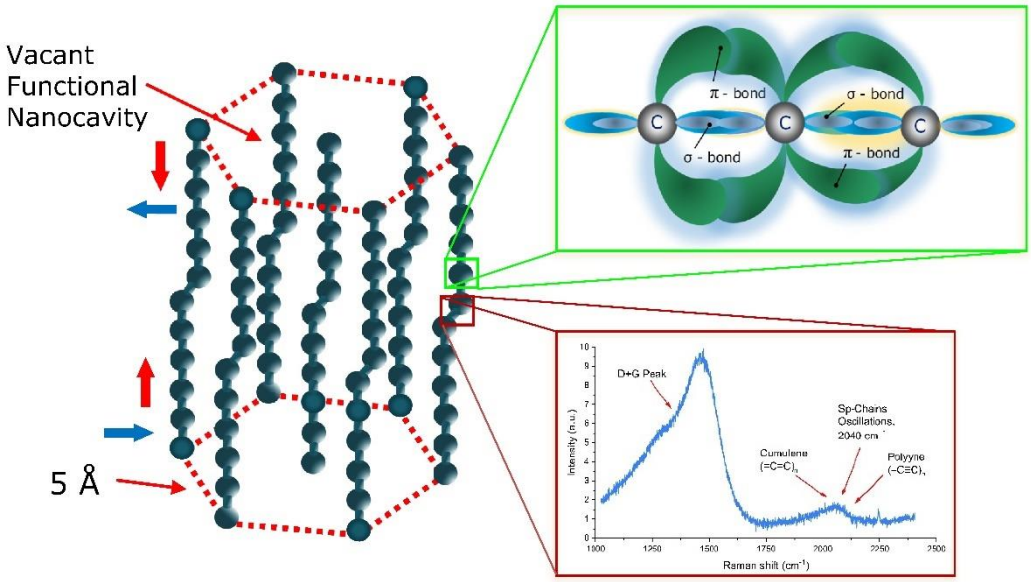

Figure 1. A diagram depicting a segment of the spatial layout within the 2D LCC matrix. This fragment serves to illustrate the organized arrangement of carbon chains, tightly packed in a parallel orientation to one another. Consequently, this arrangement gives rise to hexagonal structures that are positioned perpendicular to the surface of the substrate.

The 2D LCC nano-matrix is a structural arrangement featuring multiple empty functional nano-cavities that can be utilized for integrating atom clusters comprising various chemical elements.

The unique characteristics of this carbyne-like nanomaterial have expanded possibilities for developing advanced nanodevices with a vast array of useful applications.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Understanding the correlations between the mechanisms underlying pulsed plasma growth of carbon nanomaterials and the consequent properties holds the key to successfully implementing inverse design strategies for creating innovative and promising carbon nanomaterials.

For more than thirty years, the Materials Science Tetrahedron (MST) paradigm has been a fundamental concept in the field of materials science. It offers a holistic framework that effectively represents the complex interplay between key elements. The paradigm of nanomaterials is crucial in understanding their complexities, and it consists of four essential elements: Structure, Properties, Performance, and Processing. This framework holds immense significance in deciphering the enigma surrounding nanomaterials. A visual diagram illustrating the essential constituents of the Tetrahedron materials science framework is presented on Figure 2.

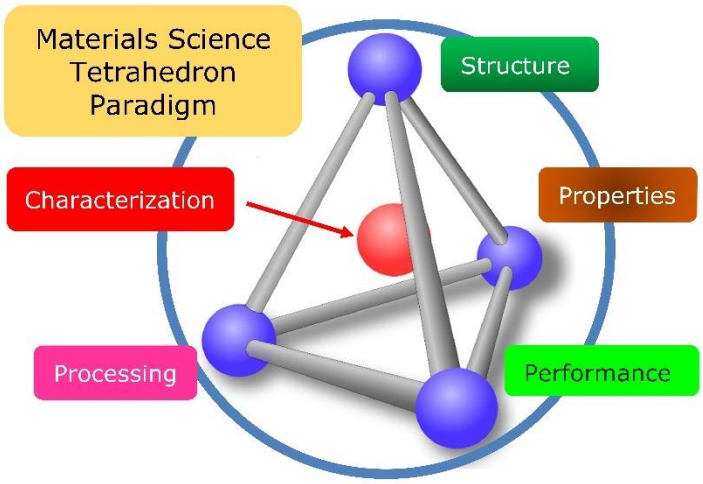

Figure 2. Visual representation highlighting the key constituents of the MST paradigm. This paradigm consists of four interconnected aspects: material structure, properties, performance characteristics, and processing techniques. Conducting a thorough analysis of these elements plays a crucial role in acquiring a deep understanding of material properties and developing innovative materials with specific attributes.

Additionally, it acknowledges the paramount importance of characterization in enabling the analysis and assessment of properties exhibited by nanomaterials.

The MST paradigm is a conceptual framework encompassing essential elements of materials science research and development. It portrays distinct dimensions of materials characterization and performance at each vertex, which work together in synergy to enhance the comprehensive comprehension of materials.

The first vertex focuses on comprehending and predicting the structure of materials at various scales, spanning from atomic arrangements to macroscopic properties. Employing techniques like crystallography, microscopy, and spectroscopy, researchers visualize and analyze the composition and arrangement of materials.

The second vertex delves into the investigation of material properties, encompassing mechanical, electrical, thermal, and optical characteristics. Through rigorous testing, researchers assess how materials perform under diverse conditions and stimuli, thereby gauging their suitability for specific applications.

The third vertex encompasses materials processing and synthesis, wherein materials are expertly designed and manipulated at the atomic and molecular levels to achieve desired properties and performance. Casting, sintering, alloying, and surface modification techniques, among others, come into play to alter the structure and composition of materials.

The fourth vertex emphasizes materials performance and application, providing insights into the diverse fields in which materials find purpose, such as electronics, energy, healthcare, aerospace, and beyond. By considering functional requirements and constraints, researchers continually strive to optimize materials to meet specific needs.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Efficiently illustrating the interplay between structure, properties, performance, processing, and characterization in the investigation of low-dimensional nanocarbons, the MST paradigm functions as a guiding framework for researchers and scientists, providing valuable direction.

The theoretical understanding of the parameters involved in the ion-stimulated pulsed plasma synthesis of low-dimensional carbon nanostructures, along with their association with physicochemical and electrophysical characteristics, remains complex due to the influence of unaccounted factors. Due to the presence of an array of atoms, external fields, ion beams, plasma, and elevated temperatures, coupled with significant durations, scientists are compelled to rely on somewhat basic assumptions.

However, understanding the immense complexity of carbon nanomaterials requires a significantly more robust approach. As a result, the development of the Digital Twin within the MST Paradigm has become imperative.

2. The Fusion of Materials Science Tetrahedron Paradigm and Deep Nanomaterials Informatics

Lately, advancements in nanotechnology have led to notable strides and an overwhelming volume of data. This emphasizes the urgent need for innovative approaches to proficiently handle, assess, and extract valuable insights from this vast pool of information.

Machine learning (ML) has recently become widely recognized and influential across various scientific disciplines, including the field of materials science, [8–12]. These approaches focused on data have greatly expedited the process of exploring and developing innovative materials. ML facilitates significant advancements by anticipating complex physical phenomena that surpass the abilities of current theoretical models.

“Material informatics” and its subcategory at the nanoscale known as “nanoinformatics” utilize extensive datasets to perform quantitative analysis on nanostructures. This facilitates an in-depth comprehension and accurate prediction of their properties, [13].

In this particular area of study, known as “nanoinformatics,” the main emphasis lies on uncovering the quantitative connections between nanostructures and their properties. The progress made in both experimental and computational techniques for characterizing nanomaterials has emphasized the significant role that “nanoinformatics” plays in attaining a comprehensive comprehension of nanomaterial behavior.

By consolidating diverse datasets into unified repositories, nanoinformatics facilitates the training of models capable of predicting the connections between features of nanomaterials and their functional behaviors. Structure-property relationship modeling is employed to intentionally design engineered nanomaterials in a reverse manner, aiming to attain specific properties that are desired.

Nanoinformatics provides the necessary framework for extracting valuable insights from diverse nanoscale data. By utilizing informatics techniques, it enhances the effectiveness and output of nano-science studies, facilitates the development of nanotechnology products, and advances the evaluation of the health and environmental effects of nano-materials.

By merging the MST paradigm with deep nanomaterials informatics, a synergistic approach is unleashed to drive the development of nanomaterials science. Through this integration, researchers can harness the complementary strengths of both frameworks, resulting in an expedited process of discovery, optimization, and design of nanomaterials. This breakthrough fosters transformative advancements in the realm of nanotechnology.

Furthermore, the amalgamation of deep nanomaterials informatics with the MST paradigm can provide valuable assistance in the exploration of materials, particularly in domains where experimental data is scarce. Through the utilization of existing databases and predictive models, researchers are empowered to make well-informed decisions regarding the compositions and structures of potential materials. This capability significantly reduces the requirement for exhaustive experimental trials, saving time and resources. Additionally, the application of deep nanomaterials informatics can enhance the efficacy of the MST paradigm by facilitating high-throughput screening of materials.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

The concept of the digital twin has evolved through extensive research conducted by scholars over the course of multiple decades. In his groundbreaking book titled "Mirror Worlds," published in 1991, David Gelernter introduced a revolutionary concept of constructing virtual representations of physical objects. Furthermore, in 2002, Dr Michael Grieves of the University of Michigan pioneered the exclusive application of digital twin technology for manufacturing purposes, introducing the concept of digital twin software. In 2010, John Vickers, a prominent figure at NASA, played a significant role in popularizing the term "digital twin." Nevertheless, it is important to acknowledge that NASA likely employed similar technologies during the 1960s for their groundbreaking Apollo missions.

A digital twin refers to a computer-generated virtual representation that encompasses the complete lifecycle of an object or system. The digital twin undergoes constant updates by utilizing real-time data. It harnesses the power of simulation, ML, and reasoning techniques to effectively assist in decision-making processes. At its core, a digital twin is a painstakingly developed virtual model that faithfully replicates and reflects the physical object or system it represents.

By enabling the generation, testing, and experimentation of hypotheses through computer modeling and simulation, digital twins serve as valuable tools in acquiring knowledge and exploring new insights. Digital twins are a robust instrument for conducting experiments, making comparisons, and unearthing innovative ideas, all without the necessity of physical implementation, [14].

Digital twins surpass traditional models and simulations through the creation of dynamic connections between the physical and digital domains, effectively harnessing the immense potential of cutting-edge technologies. Digital twins play a crucial role in expediting AI and data science endeavors, particularly within the manufacturing industry.

A digital twin incorporates three vital components: physical entities, virtual models, and interconnected data, facilitating a two-way communication channel between them.

The primary goal of digital twins is to engage in an iterative process of designing, evaluating, and refining a physical object within the virtual domain until the model reaches the desired level of performance. Once this milestone is reached, the object is considered prepared for physical implementation or for further enhancements if it already exists in the physical realm.

Considerable advancements have been achieved in facilitating a smooth integration between the MST paradigm and deep nanomaterials informatics. These advancements are the result of the innovative framework known as the digital twin of the MST paradigm.

3. Unleashing the Power of Data: A Data-Driven Nanocarbon Genome Approach

In order to effectively integrate the MST paradigm with nanoinformatics, we have devised a method centered around the notion of a digital twin for the MST paradigm. This approach takes the form of a data-driven nanocarbon genome approach (NCGA), which essentially serves as a digital twin constructed based on the principles of the MST paradigm.

The development of the data-driven NCGA is intricately connected to the inception of the "Materials Genome" concept, which took root in the United States in 2011 through the Materials Genome Initiative (MGI), [15,16]. In 2015, a collaborative effort was undertaken by scientists from both Canada and the United States, marking the inception of the Nanomaterials Genome Initiative (NMGI), [17]. The main aim of this initiative is to streamline the process of developing and incorporating highly promising nanomaterials, ensuring their rapid integration into various applications.

Materials genome technology signifies a cutting-edge approach in materials research, surpassing the conventional trial-and-error method with its advanced techniques. Materials genome technology utilizes advanced analytical software and efficient data organization methods in combination with high-productivity experimental techniques. Data is evaluated through the use of computational tools to investigate potential connections between material parameters and material properties. By adopting this approach, the process of discovering the ideal material becomes more efficient. This approach has the potential to amplify experimental efficiency, decrease costs, and minimize potential errors in experiments.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Our pioneering data-driven NCGA is designed to uncover concealed relationships between the parameters of pulsed plasma growth of carbon nanostructures and their structural, physicochemical properties, as well as performance characteristics. The adoption of the data-driven NCGA represents a revolutionary advancement in the realm of materials science and nanotechnology.

Figure 3 provides a schematic illustration demonstrating the amalgamation of the MST paradigm and deep nanomaterials informatics. This integration is facilitated through a computational "digital twin" approach termed the data-driven NCGA.

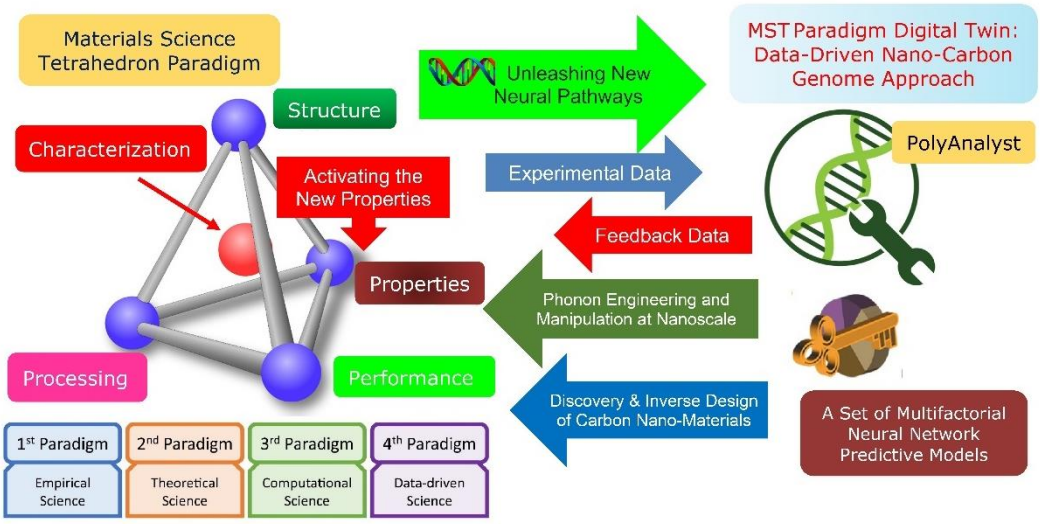

Figure 3. Schematic depiction showcasing the fusion of the MST paradigm with deep nano-materials informatics using the computational approach of "digital twin" - data-driven nanocarbon genome approach.

By applying systematic organization and analysis of data on the growth of low-dimensional nanocarbons, the data-driven NCGA harnesses artificial neural networks to process information on their growth modes, structure, as well as physical and chemical properties.

Within the NCGA framework, each specific carbon nanomaterial encompasses a distinctive signature, which represents a unique collection of properties and characteristics intricately linked together. The distinct signature inherent in each carbon nanomaterial forms the fundamental basis for its identification, classification, and subsequent application across diverse scientific and technological fields.

The data-driven NCGA comprises a collection of multifactor predictive models that not only identify dependencies and relationships within experimental data but also enable the execution of virtual experiments beyond the confines of this data space.

Utilizing these predictive models facilitates the resolution of both direct and inverse problems related to predicting the properties of nanomaterials. These models account for specific synthesis conditions and offer insights into determining the optimal parameters for achieving nanomaterials with desired characteristics. In order to establish a comprehensive system of multifactor predictive models that encompass experimental data and accommodate intricate relationships among numerous factors, deep learning neural networks (DLNN) are employed.

The data-driven NCGA distinguishes itself from the conventional notion of a physical genome as understood in biology. Conversely, it offers a conceptual framework that aids in comprehending and categorizing the extensive diversity of carbon nanomaterials. The approach utilizes data collection, analysis, and interpretation to gather information about the composition and characteristics of carbon-based nanomaterials. The data-driven NCGA can be likened to a neural network, where the connections symbolize the associations between growth factors and properties of carbon nanomaterials.

As we accumulate more data, the genome continues to expand, providing us with a deeper understanding of the vast possibilities of carbon nanomaterials.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

The true strength of the data-driven NCGA lies in its capacity to unleash the immense untapped potential of carbon nanomaterials, thereby paving the way for the creation of nanomaterials boasting entirely novel properties. The data-driven NCGA is more than a mere concept and forecasting tool; it serves as a powerful instrument for exploration, enabling us to uncover new insights and chart a path towards the future of nanotechnology.

It ushers us into an era marked by unparalleled discovery and innovation, where our comprehension of the world's most fundamental level – the nanoscale – is propelled by data. The future has arrived, and it is fueled by the power of data.

The data-driven NCGA, along with its developed methodology for implementation, acts as a catalyst in the pursuit of identifying optimal technological parameters and growth modes for carbon nanostructures. This approach enables us to growth carbon nanostructures with precise topological and physicochemical characteristics, aligning with our desired outcomes.

Through genome analysis, scientists can uncover previously undiscovered connections between the properties of nanocarbon materials.

This process enables them to identify relationships that were previously unknown, expanding our understanding of these materials. This breakthrough paves the way for the development of novel materials with customized properties to meet specific requirements, presenting a nearly boundless range of potential applications.

In order to develop and enhance the data-driven NCGA information base, extensive experimental work was conducted as part of this study. The objective of this study was to analyze how different factors affect the properties of carbon nanostructures through ion-assisted pulsed plasma growth. As a result of our findings, we have successfully identified universal key descriptors that accurately characterize the properties of nanomaterials in a clear and unambiguous manner.

The multifactor predictive models were focused on several key functions to assess the properties of the nanomaterials. These functions included:

1. I_{sp}/I_G Ratio: This ratio signifies the relative intensities observed in the Raman spectra, specifically between the peak representing sp-hybridized carbon chains and the peak denoting graphite bonds within the structure. This ratio provides a direct measure of the proportion of sp-hybridized carbon in the thin film's overall matrix of amorphous carbon;

2. Piezoelectric Surface Acoustic Waves (SAW): The frequency and shape of these waves, generated in the sample growing area, were analyzed to gain insights into the properties of the nanostructures;

3. Concentration of Methane, Hydrogen, and Argon: The relative concentration of these gases in the mixture supplied to the vacuum chamber during sample growth was studied to understand their influence on the resulting nanostructures;

4. Discharge Voltage of Carbon Plasma Source: The voltage applied to the carbon plasma source was considered as a factor that could affect the formation of the nanostructures;

5. Ion Source Energy: The energy applied to the ion source was also taken into account as an important parameter impacting the properties of the resulting nanostructures;

6. High-Frequency Electromagnetic Irradiation Parameters: The parameters of the electromagnetic irradiation used to initiate remote self-assembly of carbon nanostructures during sample growth were examined to determine their impact on the final properties of the nanostructures.

Due to the replication of universal patterns during the development of carbon nanostructures from carbon plasma at high temperatures, the shapes of these structures are greatly influenced by the vibrational state of carbon plasma molecules. Consequently, a vital aspect for assessing the attributes of these carbon nanostructures lies in analyzing the spectra of molecular vibrations. Raman vibrational spectroscopy allows for the measurement of these spectra, providing crucial and universally applicable information about the unique properties of each carbon nanomaterial. Raman spectra contain valuable quantum information that can be extracted and interpreted using sonification techniques. These techniques enable the conversion of vibrational spectra into acoustic holograms, unlocking the hidden potential of the data.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

To put it simply, the data-driven NCGA serves as a conceptual framework that effectively classifies the wide range of characteristics exhibited by nanocarbon materials. Through the collection and analysis of data, scientists can accurately anticipate the properties of novel materials, uncover uncharted correlations, and engineer materials that cater to specific requirements. The potential applications of such advancements are expansive and profound, heralding a future where the boundaries between science fiction and scientific reality blur.

4. Constructing Robust Multifactor Predictive Models

Figure 4 showcases our iterative ML workflow, which involves the seamless integration of the data-driven NCGA with the development of multifactorial neural network predictive models. The diagram illustrates the various crucial stages encompassed in the workflow.

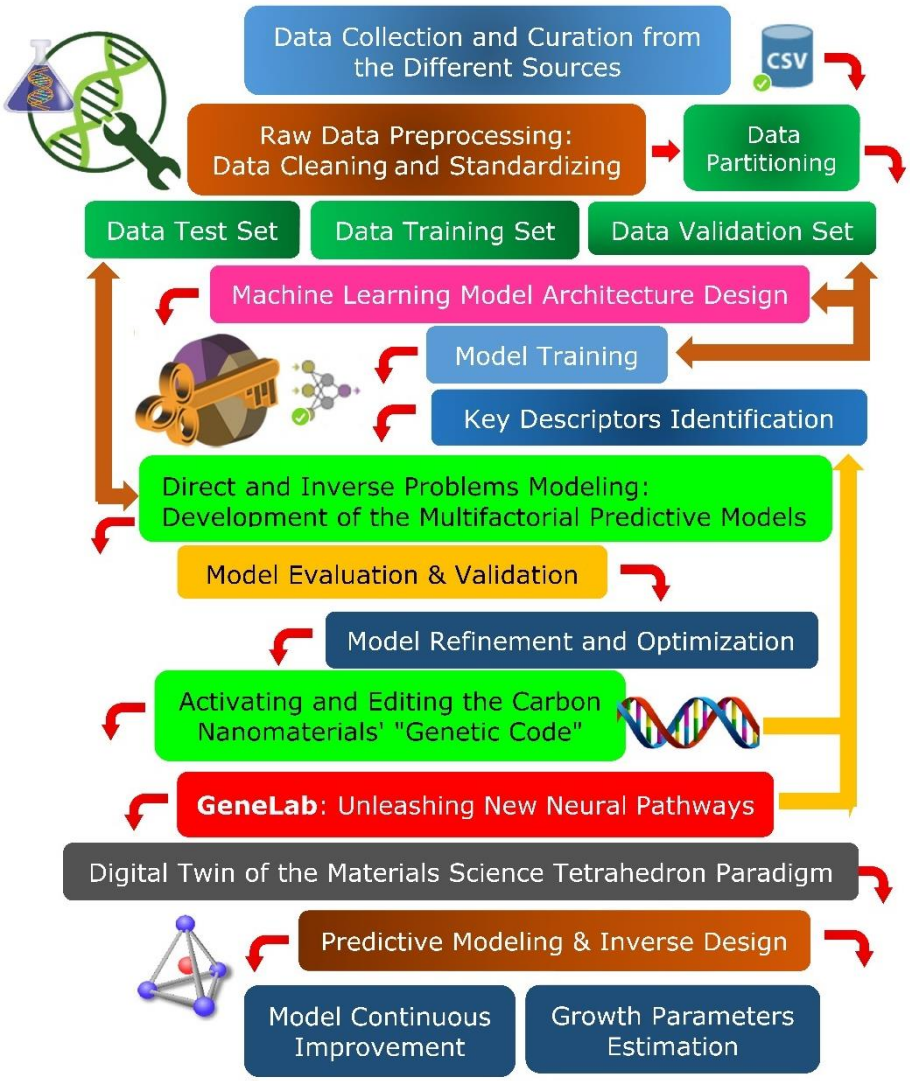

Figure 4. Schematic representation of the ML workflow employed for the development of multifactor predictive models of nanocarbon growth within the data-driven NCGA framework.

After the data cleaning stage, it is customary to partition the dataset into three separate subsets: the "Data Test Set," the "Data Training Set," and the "Data Validation Set." In the ML workflow, the models undergo training using the data training set, validation using the data validation set, and evaluation using a distinct and untouched data test set. By following this procedure, a fair and unbiased assessment of the models' performance and their capability to generalize is ensured. By evaluating the models using test datasets, we assess their predictive power and their capacity to generalize. The refinement process of the models entails adjusting hyperparameters and incorporating additional features to enhance their performance.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

To develop accurate multifactor predictive models, we leverage state-of-the-art neural network architectures that excel at capturing complex relationships between various factors and nanocarbon properties. These models efficiently capture the intricate and multifaceted relationships between synthesis conditions and the properties of nanocarbon, effectively allowing for precise predictions and providing valuable insights. The models undergo training using preprocessed data and are subsequently validated to evaluate their performance.

DLNN can be effectively leveraged to develop multifactor predictive models for addressing direct and inverse problems in carbon nanomaterial growth via pulsed plasma techniques. DLNN effectively uncover and comprehend the intricate relationships between various growth parameters and the resultant properties of nanocarbons. By training these models on diverse datasets encompassing a wide range of input parameters and corresponding output characteristics, they provide valuable insights into the dependencies and associations between the growth parameters and nanomaterial properties.

To address direct problems involving the correlation between the topological and physicochemical properties of carbon nanomaterials and pulsed plasma growth parameters, deep learning models can undergo training on datasets comprising input parameters (such as plasma power, gas composition, and growth time) and their corresponding outputs (like material structure and electrical properties). Subsequently, this trained model can predict the characteristics of carbon nanomaterials based on the provided growth parameters, allowing for interpolation of data within the range covered by the training dataset.

Inverse problems refer to the challenge of identifying the growth parameters necessary to achieve specific topological, physicochemical, and operational traits in carbon nanomaterials. To address this, a framework utilizing deep learning models can be employed, wherein the desired output characteristics serve as input for the model to learn and predict the associated growth parameters.

The effectiveness of the deep learning model in resolving inverse problems is contingent on several key factors. These include access to a comprehensive and representative dataset, implementation of suitable preprocessing techniques, and thoughtful selection of architectural choices. Additionally, the incorporation of domain knowledge and experimental validation plays a pivotal role in ensuring the accuracy and practicality of the predicted growth parameters.

Training a model to tackle inverse problems can pose greater challenges as it entails acquiring an understanding of the intricate relationship between desired output characteristics and the growth parameters. This often calls for a more extensive and diverse dataset to encompass the entire range of possible solutions.

Below is a detailed description of the general structure of DLNN used to develop multifactor predictive models for carbon nanomaterials obtained through pulsed plasma growth methods:

1. Input Layer:

- In a neural network, the input layer receives the input features or factors that impact the growth of carbon nanomaterials via pulsed plasma growth;

- The size of the input layer varies based on the quantity of input factors, including plasma power, precursor concentration, growth time, reactor temperature, and any other pertinent parameters.

2. Hidden Layers:

- DLNN generally comprise several hidden layers, which apply nonlinear transformations to the input data;

- Experimentation and validation are necessary to determine the optimal number of hidden layers and the number of neurons within each layer. Increasing the number of layers enables the network to learn intricate patterns and relationships within the data;

- Each hidden layer incorporates an activation function to introduce nonlinearity and calculate the output of each neuron. Commonly used activation functions include ReLU (Rectified Linear Unit), sigmoid, and hyperbolic tangent.

3. Feature Extraction:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

- The hidden layers of the neural network carry out the extraction of pertinent features from the input factors, capturing intricate relationships and patterns;

- If appropriate, several techniques such as convolutional layers or recurrent layers can be integrated into the network to effectively handle specific data structures or temporal dependencies.

4. Output Layer:

- The output layer of the network generates predictions or estimates concerning the behavior or properties of carbon nanomaterials acquired through pulsed plasma growth;

- The configuration of the output layer varies depending on the specific task at hand. For instance, it can predict material characteristics such as morphology, size, conductivity, or any other desired property of the nanomaterials;

- The selection of the activation function within the output layer is contingent on the nature of the predicted property. A linear activation function is suitable for continuous values, while sigmoid or softmax activation functions are more appropriate for classification tasks.

5. Loss Function:

- The selection of the loss function is contingent on the particular problem being tackled. Mean squared error (MSE) or mean absolute error (MAE) are commonly used loss functions for regression tasks;

- In classification problems, the cross-entropy loss function is frequently employed, with suitable adjustments made for multi-class classification if necessary;

- It is also feasible to tailor the loss function to accommodate additional penalties or constraints that are specific to the problem being solved.

6. Training and Optimization:

- The network undergoes training using backpropagation, through which the weights and biases are adjusted incrementally to minimize the selected loss function;

- Optimization techniques such as stochastic gradient descent (SGD) or more advanced optimizers are utilized to update the network's parameters;

- To ensure proper training, the dataset is usually divided into training, validation, and testing sets. The validation set aids in preventing overfitting by monitoring the model's performance throughout training, while the testing set is employed to assess the final model's ability to generalize.

7. Model Evaluation:

- The model's performance can be evaluated using various metrics, such as mean absolute error, mean squared error, accuracy, precision, recall, or F1 score, depending on the specific problem at hand;

- To enhance the robustness and reliability of the model, additional techniques like cross-validation can be employed.

It is crucial to highlight that the structure and design choices of the DLNN can be tailored and improved according to the unique demands and characteristics of carbon nanomaterial growth through pulsed plasma techniques. The selection of layers, activation functions, and other hyperparameters can be determined through experimentation and thorough data analysis.

The iterative nature of our workflow facilitates ongoing enhancements and adjustments, ensuring continuous improvements and adaptability. By integrating new data, regularly updating models, and staying abreast of evolving knowledge, we can guarantee improved accuracy and relevance as time progresses.

In summary, our iterative ML workflow strives to offer precise predictions, valuable insights, and accessible tools for research in nanocarbon growth. By doing so, we aim to contribute to significant advancements in the field.

In our research endeavors, we employ cutting-edge predictive modeling techniques to obtain valuable insights and achieve precise predictions. To accomplish this, we leverage the robust analytical platform known as PolyAnalyst, provided by Megaputer Intelligence, [18].

PolyAnalyst is an advanced software platform that offers a wide range of tools and functionalities for comprehensive data analysis, feature extraction, and predictive modeling. It

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

facilitates the development of multifactor predictive models, which are neural networks that incorporate multiple input factors.

PolyAnalyst provides support for visual programming, enabling the implementation of data collection, processing, analysis, and modeling tasks. All operations are visually represented as an interactive script or directed graph within PolyAnalyst, with nodes representing different functions and arcs illustrating the flow of data.

Figure 5 presents an example of an interactive analytical script showcasing the extensive capabilities of data processing, analysis, and modeling offered by the PolyAnalyst analytical platform.

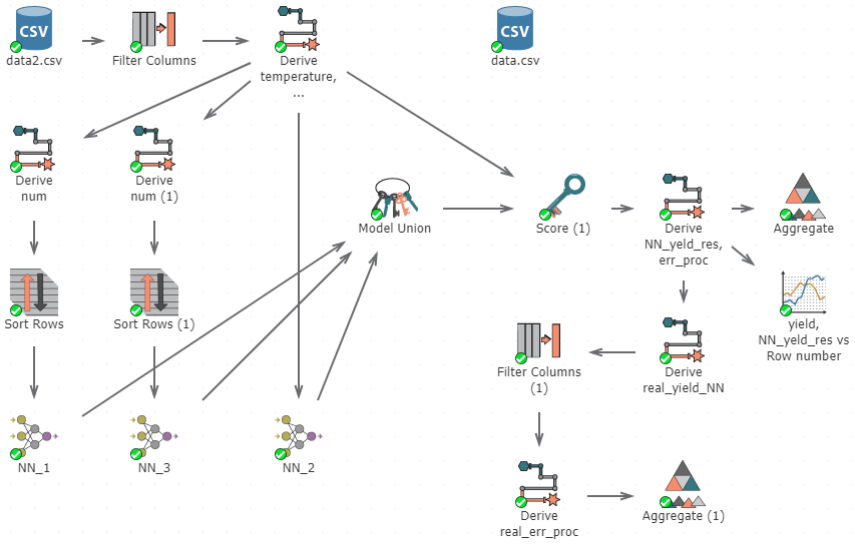

Figure 5. An example of an interactive analytical script demonstrating the comprehensive functions of data processing, analysis, and modeling within the PolyAnalyst analytical platform.

Analytical scenarios are constructed by linking functional nodes, where each node represents a specific data operation such as loading, transforming, visualizing, or implementing ML algorithms. We have the ability to define input factors, specify desired output, and employ a range of algorithms for model training and optimization. PolyAnalyst seamlessly integrates with various data sources, enabling the combination of structured and unstructured data for comprehensive analysis. This approach guarantees a more holistic comprehension of the data and augments the quality of our models.

The platform offers a user-friendly interface for constructing neural network predictive models with multiple factors, making it intuitive and easy to use. PolyAnalyst integrates techniques that enhance the interpretability and explainability of the models, ensuring a better understanding of the underlying factors influencing the predictions.

This enables us to comprehend the factors that influence predictions and acquire valuable insights into the underlying relationships embedded within the data.

5. Identifying Universal Descriptors for Nanocarbons using Machine Learning

Through our data-driven study of nanocarbons using NCGA, a significant outcome has been the identification of universal descriptors that hold significant value. The utilization of these descriptors is crucial as they guide the inverse design process and optimize the synthesis parameters, enabling the attainment of desired properties in nanocarbons. Universal descriptors are fundamental parameters or characteristics that exert a significant influence on the properties of nanomaterials, irrespective of their distinct structures or synthesis methods. Through the analysis of the interconnections between growth parameters and resulting properties, we are able to discover concealed relationships between growth and properties, thereby identifying these universal descriptors.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Descriptors serve as an information resource equipped with predictive capabilities, [19,20]. The identification of universal descriptors empowers us to streamline the design process, enhancing its efficiency and effectiveness. This process encompasses in-depth analysis and the application of data mining techniques to the collected and curated data on nanocarbons. The progress in ML algorithms and statistical analysis enables us to uncover previously undiscovered relationships between growth parameters and the resulting properties of nanocarbons. Through the application of these techniques, we can extract valuable insights from the data, enabling the identification of crucial factors that exert a substantial impact on the properties of nanocarbons. These factors include growth temperature, precursor concentration, growth time, catalyst type, and other variables that influence the structure and performance of the nanocarbons. The quantity and quality of the identified descriptors directly influence the accuracy and quality of the predicted results. By broadening the scope of descriptors, we can generate a synergy effect, leading to the discovery of fresh avenues for synthesizing nanocarbons with specific desired characteristics. Figure 6 visually depicts the data-driven process of observing and modifying key descriptors and connections. These crucial elements are seamlessly integrated into our NCGA for inverse design.

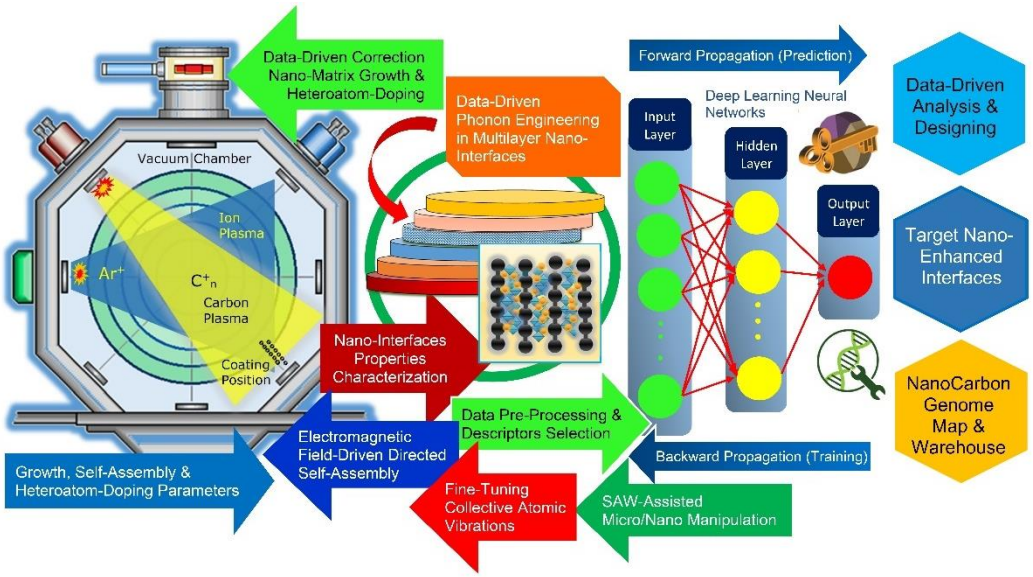

Figure 6. Visualization of the data-driven process involving the observation and modification of key descriptors and connections. These essential elements are seamlessly integrated into NCGA for inverse design.

The identification of universal descriptors proves to be immensely valuable in guiding the reverse design of nanocarbons. These descriptors function as indispensable tools in fine-tuning growth conditions and attaining the desired properties in the final nanocarbons. By leveraging these descriptors, we can effectively minimize the time and resources invested in trial-and-error experiments. Furthermore, uncovering universal descriptors grants us an in-depth comprehension of the fundamental principles that determine the correlation between growth and properties in nanocarbons. This knowledge empowers us to uncover fundamental insights into the behavior and correlation between structure and properties in nanocarbons.

By incorporating these universal descriptors into our predictive modeling framework, we can significantly bolster the precision and efficacy of the inverse design process.

In summary, the identification of universal descriptors within our data-driven NCGA marks a critical stride in unlocking the true potential of nanocarbons. By understanding the crucial parameters that influence the properties of nanocarbons, we can accelerate the design process, optimize synthesis conditions, and attain customized properties in an efficient manner.

6. Exploring the "Genetic Code" in a Nanocarbon Genome Approach

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

The establishment of a comprehensive set of data-driven neural pathways is highly dependent on the identification of all the relationships that exist in the experimental data between the parameters of pulse-plasma growth in carbon nanostructures and their complete spectrum of properties. The establishment of these relationships constitutes a critical component in the construction of the NCGA.

Through the manipulation of synthesis variables, we have the ability to reconfigure and fine-tune this circuitry, creating novel pathways while also modifying the pre-existing ones. The adjustment of neural circuitry facilitates improved regulation of material characteristics and capabilities, thereby expanding the potential of the nanocarbon genome through the implementation of redesigned genomic circuitry.

The NCGA employs distinctive signatures or "genetic codes" to characterize nanocarbons, highlighting their individuality. The "genetic code" symbolizes a connected collection of distinct properties and characteristics that define the unique nature of nanocarbons. The data-driven nanocarbon genome unveils its "genetic code" through a complex network of connections that precisely map synthesis parameters to the resulting properties. This intricate circuitry provides a comprehensive understanding of how specific synthesis conditions influence the characteristics exhibited by nanocarbon materials.

This conditional concept encapsulates the fundamental connections between different factors, including structure, composition, size, and processing conditions, and their direct impact on the physical properties exhibited by nanocarbons. By comprehending these relationships, we gain a deeper understanding of how these various factors collectively shape the characteristics of nanocarbon materials.

By harnessing the power of the conditional concept embodied in the "genetic code," we can effectively capture the intricate patterns and correlations that exist within vast amounts of data pertaining to the properties of nanocarbons. This allows us to derive valuable insights and discern meaningful relationships that may otherwise remain hidden. As we delve into the analysis and comprehension of the interconnections among diverse properties, we gain the ability to develop models and algorithms capable of predicting or inferring properties based on the unique "genetic code" of a specific carbon nanomaterial. This implies that by deciphering the underlying relationships, we can construct powerful tools that allow us to anticipate and extrapolate properties with a high degree of accuracy. With the utilization of this capability, we can effectively navigate and make the most of the vast array of carbon nanomaterials and their associated properties. This not only simplifies the process of exploration but also empowers us to utilize their potential in developing groundbreaking applications and advancements across diverse technological domains.

By dissecting this "genetic code," researchers can unlock valuable insights into how specific combinations of these factors dictate the distinctive characteristics of carbon nanomaterials. The concept of the "genetic code" is essential in understanding the intricate balance and interdependence among various variables, contributing to their impact on the properties of nanocarbons. By deciphering this "genetic code," we have the ability to design nanocarbons with specific characteristics, anticipate and control their properties to meet particular application requirements.

7. Activating the Nanocarbon's "Genetic Code" through Unleashing Novel Neural Pathways

The data-driven nanocarbon genome can be conceptualized as a neural network, with its connections symbolizing the interconnectedness between growth factors and the properties of the nanocarbon. The notion of activating the "genetic code" of nanocarbon can also be perceived as stimulating fresh neural pathways within this network.

Through proactive exploration and establishment of novel neural pathways within the framework of the nanocarbon's "genetic code," we can unravel the intricate connections between synthesis parameters, nanocarbon structures, and the resulting properties. This comprehension holds utmost importance in fine-tuning the synthesis process to achieve specific structural characteristics and, consequently, the desired properties of low-dimensional nanocarbons.

Traditional experimental methods often do not possess the capacity to systematically generate a comprehensive map of the complex connections between adjustable synthesis variables and the wide

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

array of resulting nanostructures. Nevertheless, by activating novel neural pathways, we can unveil hidden relationships and connections that were previously undisclosed or disregarded.

The process of forging innovative linkages between synthesis mechanisms and resulting properties through physics-based approaches possesses immense potential to revolutionize the field of inverse nanocarbon design. This transformative approach has the capability to completely reshape the way we design nanocarbon materials. These novel pathways serve as gateways, exposing previously concealed connections between structure and properties. By undertaking this exploration, we discover concealed neural pathways that enable the emergence of exceptional properties in carbon nanostructures.

There are several advantages associated with the activation of new neural connections in the NCGA:

- Inclusion of additional parameters: Introducing new neural pathways enables the consideration of additional parameters that influence the properties of carbon nanomaterials. This inclusion allows for a more precise characterization of properties and growth processes by incorporating relationships between chemical composition, synthesis temperature, and nanomaterial structure;

- Detection of hidden dependencies: The establishment of new neural pathways enables the identification of non-obvious relationships and dependencies among carbon nanomaterial parameters. This discovery may lead to the exploration of new properties and the utilization of unpredictable parameter combinations to create nanomaterials with desired characteristics, thereby enhancing the properties of resulting carbon nanomaterials;

- Enhanced prediction accuracy: By creating new neural pathways, it becomes possible to improve the accuracy of property predictions for carbon nanomaterials. Incorporating more information and relationships significantly enhances modeling and prediction precision, facilitating the determination of optimal parameters for achieving desired nanomaterial properties;

- Advancement in reverse engineering: The creation of new neural pathways within the NCGA facilitates a more accurate and efficient design process for carbon nanomaterials. By leveraging additional information and relationships, the reverse engineering approach becomes predictive and informative, thereby enhancing the quality and efficiency of producing carbon nanomaterials with desired properties.

In summary, the activation of new neural pathways in the data-driven NCGA enables the consideration of a broader range of parameters, the detection of previously unseen dependencies, the enhancement of prediction accuracy, and an increase in reverse engineering efficiency. These advancements allow for the creation of carbon nanomaterials with specific sets of properties in a more accurate and efficient manner. Embarking on uncharted synthesis pathways unveils new neural routes within the framework of the data-driven nanocarbon genome. This exploration allows for navigation through previously inaccessible realms in the data-driven genome, leading to the discovery of fresh connections between growth conditions and nanomaterial properties. This unearths vast potential for precise engineering of tailored nanocarbon materials.

8. Emerging Technologies and Approaches for Activating Novel Neural Pathways

Our ability to activate new neural pathways significantly amplifies the predictive potential of the data-driven nanocarbon genome. This remarkable advancement empowers us to engineer nanocarbons with desired properties at a heightened degree, ultimately expediting the process of reverse engineering these nanomaterials.

Through our comprehensive research, we extensively explored the impacts of various factors, including heteroatom doping, external high-frequency electromagnetic fields, excitation of piezoelectric SAW, plasma parameters, and energy injection. These factors possess a remarkable ability to establish new original links between growth mechanisms and the resulting properties of carbon nanomaterials. As a result, this clears the path for the development of physics-guided neural pathways that greatly amplify predictive abilities and unlock vast potential in the realm of inverse design for nanocarbons.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Heteroatom doping refers to the incorporation of foreign atoms into the nanocarbon lattice structure, bringing about improved properties and functionalities in the nanomaterial. Through this manipulation, nanocarbons undergo substantial changes in their electronic and chemical properties, ultimately allowing for the development of innovative functionalities. By examining the impacts of distinct dopants and their concentrations on both morphology and properties, we can enhance our capability to design nanocarbons that possess customized characteristics.

External fields, specifically electric or magnetic fields, possess the ability to manipulate and regulate the growth and organization of carbon nanostructures. Gaining a comprehensive understanding of the interaction between such fields and growth parameters empowers us to control the shape, composition, and unique characteristics of nanocarbon materials. This presents captivating opportunities for tailoring nanocarbon materials with specific attributes, such as enhanced conductivity or superior optical properties, through precise engineering. The application of the Teslaphoresis force field provides a noteworthy example, as it allows for the direct self-assembly of nanocarbon structures with low dimensionality, [21].

As the structural, physicochemical properties, and functional characteristics of nanomaterials originate at the atomic scale, the crucial factor for activating new neural pathways in the NCGA lies in the programmable manipulation of collective atomic vibrations within multilayer nano-systems, [22]. Specifically, when considering nanolayers with sizes below 10-20 nm, the vibrations occurring in the outermost atomic layers are notably substantial and, consequently, they wield a significant influence on the development of their properties. By precisely adjusting the layer thickness and optimizing the geometry of the multilayer structure, a remarkable capability emerges to generate standing phonon waves. This method provides a significant amplification of the amplitudes of atomic vibrations, thereby optimizing and enhancing the desired characteristics to their maximum potential. The growth of multilayer nano-interfaces offers an extraordinary opportunity to initiate and precisely control the captivating phenomenon of self-synchronization in collective atomic vibrations. This capability directly impacts the nano-architecture and properties of nano-systems, empowering precise programming and manipulation with remarkable control. Figure 7 provides a schematic representation of an experimental multilayer thin film nano-system, elegantly constructed using 2D LCC as the foundational framework.

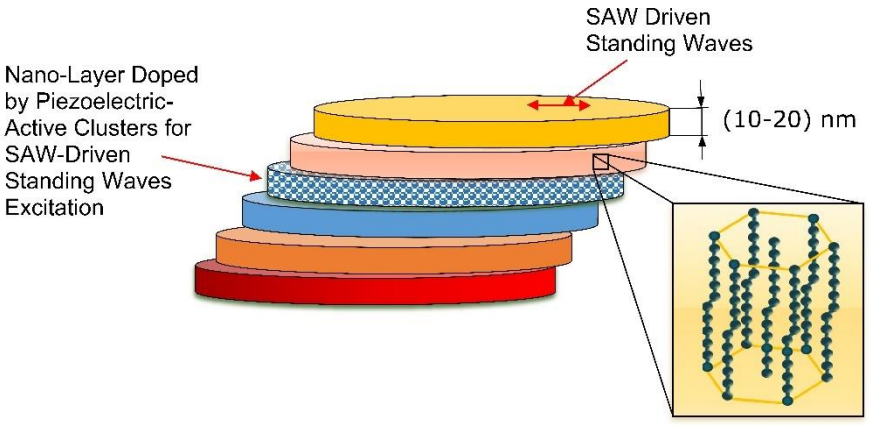

Figure 7. Schematic illustration of an experimental multilayer thin film nano-system constructed with a foundation of 2D LCC.

The utilization of piezoelectric SAWs can bring about alterations at the atomic level in the developing nanocarbon thin film, thereby modifying its phonon properties and vibrational characteristics. These transformations can be detected by observing shifts and variations in the response of Raman spectroscopy. SAWs refer to mechanical vibrations with high frequencies that propagate along the surface of a developing nanomaterial. Usually, these waves are created by inducing an oscillating electric signal onto a piezoelectric substrate. As the SAW travels through the developing nanocarbon film, it engages with the carbon atoms within the lattice structure of the film.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

This interaction induces the atoms to vibrate and oscillate at the same frequency as the surface wave. Essentially, the acoustic wave transfers a portion of its vibrational energy to the atoms within the nanofilm.

The transfer of vibrational energy from the surface wave propagating through the nanocarbon film disturbs and alters the phonon modes associated with the lattice atoms. Consequently, these modifications result in detectable changes in the Raman spectral response, providing insight into the growth process of nanocarbon. Modifications induced by the SAW on the phonon modes of the nanocarbon film result in detectable variations in the ensuing Raman spectra. The shift in frequency of the Raman peaks can be observed as a consequence of the modified phonon energies. Additionally, the intensities of the peaks can vary when different phonon states are excited. The interaction of surface waves can give rise to entirely novel peaks, which correspond to the activation of new phonon modes. Figure 8 visually illustrates the effective implementation of the SAW-based method, showcasing precise alterations made at the atomic level while nurturing the development of the nanocarbon thin film.

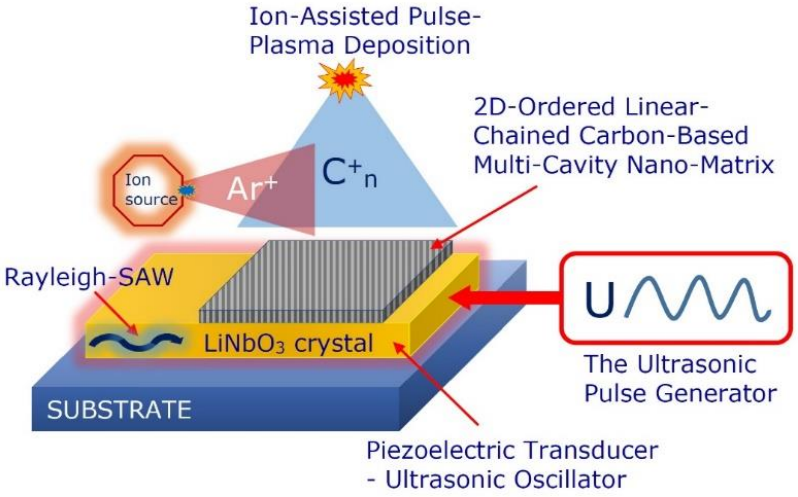

Figure 8. Illustration showcasing the effective implementation of the SAW-based technique to achieve accurate atomic-level modifications during the growth process of the nanocarbon thin film.

Another instance exemplifying the alteration of the "genetic code" responsible for a particular physical attribute in nanocarbons is the excitation of new nano-oscillatory systems. These nano-oscillatory systems serves as the source for excitation of the nano-patterns. Figure 9 showcases self-organized excitation patterns observed in amorphous carbon nitride (a-C:N) film samples grown at different ion assistance energies. Transmission electron microscopy analysis reveals that the structure of samples deposited without ion assistance appears uniform, whereas the structure of ion-assisted samples exhibits heterogeneity. The provided fragments depict sections of the self-organized patterns, each corresponding to different ion assistance energies: (A) 0 eV, (B) 200 eV, (C) 300 eV, (D) 400 eV, (E) 600 eV, and (F) 800 eV, [23].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

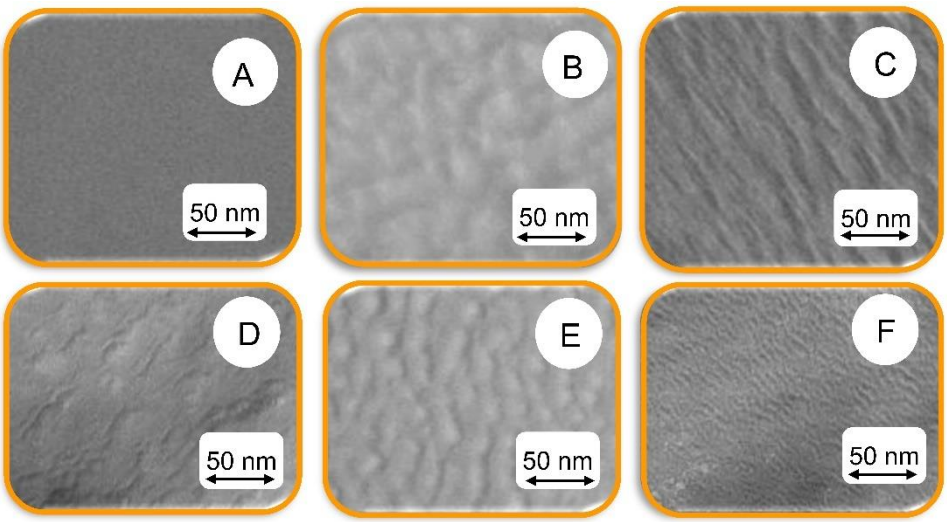

Figure 9. Self-organized excitation patterns observed in amorphous carbon nitride (a-C:N) film samples grown at varying ion assistance energies, [23]. Transmission electron microscopy reveals a homogeneous structure for samples without ion assistance, while ion-assisted samples exhibit heterogeneous structures.

9. The Role of the Data-Driven Nanocarbon Genome in the Inverse Designing of New Low-Dimensional Nanocarbons

The data-driven nanocarbon genome plays a pivotal role in enabling the inverse design of novel low-dimensional carbon nanostructures such as graphene, carbon nanotubes, and carbyne chains.

The genome encompasses a vast repository of structured data detailing the relationships between growth parameters, processing techniques, and resultant properties for a wide range of nanocarbon allotropes. By training sophisticated ML models on this aggregated data, the genome allows for the establishment of predictive correlations between synthesis conditions and material traits.

Leveraging these predictive models facilitates the resolution of inverse problems, wherein target topological features and functional characteristics are defined, and the models suggest optimal paths for attaining those properties by working backwards from the desired outputs.

For instance, the models can propose precursor compositions, growth temperatures, doping levels, and post-synthesis processing steps required to obtain graphene sheets or carbon nanotubes with specific conductivity, strength, surface area and other precisely tailored attributes.

The models account for intricate quantum-mechanical phenomena governing carbon bonding, lattice formation, allotrope transformation, and property manifestation. The neural network architectures continuously enhance their understanding of these complex mechanisms through recursive model training.

By effectively navigating the multi-dimensional parameter space and narrowing down ideal combinations of controllable factors, the carbon genome allows researchers to engineer nanocarbons de novo rather than rely solely on trial-and-error.

The genome-driven inverse design of low-dimensional carbons also suggests unfamiliar processing routes that may catalyze unconventional nanostructures with exceptional qualities. Exploring these new synthesis avenues expands the experimental search space.

Overall, the predictive power and expanding knowledge graph of the data-driven nanocarbon genome can tremendously accelerate the design, discovery and development of graphene, nanotubes, and carbyne chains with user-defined architectural and functional targets.

10. Discovering Unexplored Realms of Practical Applications for Carbyne-Enriched Nanocarbons.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

The groundbreaking activation of novel neural pathways through the utilization of the data-driven nanocarbon genome has revolutionized and expanded the horizons of techniques for growth and tailoring low-dimensional nanocarbon allotropes.

By unlocking this extraordinary potential, previously concealed properties and untapped benefits of these nanomaterials are now unveiled, ushering in a new wave of practical applications across a myriad of scientific and technological arenas. The exploration of uncharted territories and unprecedented realms of possibilities for carbyne-enriched nanocarbons has become an imminent and transformative reality.

Here are a few exciting applications that could be made possible by activating new neural pathways in the data-driven nanocarbon genome:

- Development of cutting-edge multifunctional nano-coatings and nano-barriers (referred to as “Space Skin”) designed specifically for small-sized spacecraft in low Earth orbits and spacecraft structures. These advanced nano-coatings offer a range of benefits including protection against radiation and thermal damage, prevention of material erosion caused by atomic oxygen, as well as energy harvesting capabilities for future use;

- Development of groundbreaking metamaterials with the capability to harness and harvest energy from the surrounding outer space, offering a sustainable and alternative power source for nanosatellites;

- Development of state-of-the-art ferromagnetic fluids specifically tailored for next-generation space applications. These advanced ferrofluids exhibit superior thermal conductivity and can be externally controlled and precisely adjusted in terms of their heat transfer coefficient. Consequently, they have the potential to compete with conventional heat transfer solutions. Originating from NASA, ferrofluids were first invented by Steve Papell in 1963 with the aim of creating a liquid rocket propellant that could be magnetically attracted to a pump inlet under zero-gravity conditions;

- Development of cutting-edge systems capable of efficiently converting thermal energy into electrical energy, offering high power density and extended service life. These advancements pave the way for meeting the urgent demand to transition towards environmentally friendly energy sources and align with national carbon neutralization strategies;

- Development of promising nano-coatings, enabling the conversion of cosmic radiation energy into electricity by precisely tuning the vibration properties of multilayer carbyne-enriched. These innovative coatings hold immense potential for efficient energy harvesting from cosmic radiation;

- Development of promising energy storage systems with exceptional energy efficiency;

- Development of the revolutionary “Brain Neural Matrix” capable of capturing and recording electromagnetic radiation emitted by the brain, thereby enabling precise, long-term monitoring of brain activity;

- Synthesis of a new generation of nano-catalysts for use in 3D printing technologies of catalytic elements that convert hydrogen peroxide into a gaseous phase used as rocket fuel. This technology is specifically employed in the production of the Mk II Aurora space plane by Dawn Aerospace, whose engine relies on hydrogen peroxide;

- Development of an advanced generation of nano-liquids designed for efficient absorption of solar energy, enabling its direct conversion into electricity and offering a groundbreaking solution for sustainable energy generation;

- Development of promising smart nanomaterials for biomedical applications, including drug delivery, bioimaging, biosensors, and tissue engineering;

- Development of nanocarriers for precision drug delivery as well as scaffolds for tissue regeneration;

- Development of promising thermionic cooling systems, which effectively cool down the surfaces of high-speed vehicles. This technology has the ability to absorb heat fluxes of up to 25 MW/m^2, which is significantly higher than the limits of existing thermal protection systems. These systems are typically designed to handle fluxes of 1-3 MW/m^2 at temperatures up to 2000 K;

- Development of highly efficient solar photovoltaic cells, new generations of displays, light-emitting devices, fast X-ray detectors, X-ray tubes and night vision devices;

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

- Development multi-purpose nano-sensor platforms for biomedical applications and space missions utilizing the additive manufacturing of low-dimensional nanocarbons;

- Development of nano-additives incorporating catalytic agents for enhancing the energy properties of advanced energetic materials, including gunpowder, solid propellants, and pyrotechnic compositions;

- Development of hybrid nano-additives that can enhance the energy efficiency of eco-friendly liquid rocket fuel based on hydroxylammonium nitrate (HAN), which is used in the attitude control engines of small spacecraft;

- Development of a new generation of highly sensitive nano-sensor architectures for environmental monitoring and important medical applications, such as DNA lysis and bacterial strain detection. Such nano-biosensors may also be useful for detecting contaminants in cells surrounding implants installed in various human organs;

- Development of hybrid carbon fibers, which find applications in various areas such as bulletproof vests, structural components of aircraft and missiles, rocket engines, as well as electrically heated clothing;

- Development of a cutting-edge generation of radioisotope thermoelectric generators (RTGs). These RTGs have the ability to harness the thermal energy released from the natural decay of radioactive isotopes, converting it into usable electricity. During prolonged space missions, RTGs serve as a crucial and primary power source when the utilization of solar panels is impractical or not feasible;

- Development of adaptive nano-membrane technologies. Nano-membranes, also known as molecular sieves, are capable of efficiently separating multicomponent media, which is in high demand across industries such as energy, chemical, food, oil refining, medicine, agriculture, and waste processing. These active nano-membranes possess a programmable structure that allows them to continually adjust their functional parameters to target specific molecules. This is achieved by manipulating the electric field parameters within the pores of the membrane;

- Development of nanoscale molecular devices that can both record and manipulate the bioelectric field of cells without causing harm to surrounding cells and tissues. This advancement creates exciting new possibilities for fundamental research. Bioelectricity, which refers to the electrical current that flows between cells, plays a crucial role in our ability to think, speak, and move. The ability to record and modulate the bioelectric fields of cells and tissues is of significant importance in promoting wound healing and potentially combating diseases such as cancer and heart disease.

The mentioned applications of carbyne-like nanomaterials represent only a fraction of the vast potential awaiting discovery through activating novel neural connections in the data-driven genome of these innovative nanomaterials.

11. Discussion

Elucidating the intricate growth-property interconnections of nanocarbons is pivotal for integrating these versatile materials into disruptive technologies. By mapping the complex relationships through a responsive data-driven genome, researchers gain a roadmap to engineer topological and functional traits precisely by design.

The proposed inverse genome paradigm is a breakthrough in navigating the vast nanocarbon possibility space. Conventional trial-and-error approaches struggle with computational bottlenecks and experimental barriers. By contrast, the genome leverages AI to suggest optimal synthesis routes to desired nanoarchitectures.

The work also offers an advanced toolkit of synergistic effects, nanointerface engineering, plasma control and other techniques to actualize the genome's models for industrial scaling. These physics-guided pathways overcome previous technical limitations in manipulating nanocarbon assembly.

Uncovering new neural connections expands the genome's scope for property predictions. Researchers can anticipate behaviors and morphologies early across more diverse materials.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Streamlining the discovery workflow will catalyze innovation cycles to develop game-changing nanoproducts.

The proposed multifactor predictive models provide researchers and engineers an efficient pathway to resolve both direct and inverse problems in nanocarbon engineering. The models can rapidly suggest optimal combinations of tunable parameters like precursor composition, temperature, pressure etc. to obtain target traits. This accelerates discovery and development of innovative nanocarbons for cutting-edge technologies.

In toto, by unlocking nano-relationships through an iterative data-powered genome, customizable carbons can transform applications from electronics to biomedicine. Seamless cyber-physical integration based on the approach could automate industrial manufacturing through real-time adaptive learning.

As neural horizons broaden, so too will the realms of utility for precisely tailored nanocarbons. Augmenting fundamental knowledge of these wondrous materials is key to driving technologies that shape our collective future. The flexible data-driven genome provides that navigation chart in multidimensions.

The NCGA is capable of forming the intelligent core of cyber-physical systems for Industry 5.0, enabling automated synthesis of tailored low-dimensional nanocarbons through real-time monitoring, adaptive optimization and feedback. The results obtained make it possible to create a prototype of an automated cyber-physical laboratory for the accelerated design and growth of nano-systems with specified characteristics for use in specific practical systems.

12. Conclusions

Through this groundbreaking research, the incredible potential of the data-driven nanocarbon genome has been revealed, leading to a revolutionary progress in the precise design of low-dimensional carbon nanostructures using inverse methodologies.

This groundbreaking research establishes a robust framework for data-driven discovery in nanocarbon engineering, leveraging the power of artificial intelligence to unlock the limitless potential of these materials.

As we continue to incorporate more data, explore new perspectives, and activate hidden neural connections within the nanocarbon genome, a new era of precision tailoring awaits, where the possibilities for nanocarbons are truly boundless.

The data-driven nanocarbon genome transcends its role as a predictive tool, evolving into a powerful instrument of discovery. It propels us towards a future where carbon nanomaterials are meticulously crafted and intricately tailored for precise applications, opening up a new era of engineering possibilities.

Author Contributions: Conceptualization, A.L.; methodology, A.L. and O.G.; validation, A.L. and O.G.; formal analysis, A.L. and O.G.; investigation, A.L. and O.G.; resources, A.L. and O.G.; data curation, A.L. and O.G.; writing—original draft preparation, A.L.; writing—review and editing, A.L.; visualization, A.L.; supervision, A.L.; project administration, A.L. and O.G.; funding acquisition, A.L. and O.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research work is jointly supported and funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) and the Russian Foundation for Basic Research (RFBR) - Russian Center of Scientific Information (RCSI) according to the research project № 20-58-46014.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The publication cost of this paper was covered by funds from the Russian Foundation for Basic Research (RFBR) - Russian Center of Scientific Information (RCSI) according to the research project No. 20-58-46014.

Conflicts of Interest: The authors declare no conflict of interest.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
CAW	Carbon-atom wire
DFT	Density functional theory
DLNN	Deep learning neural networks
DNN	Deep neural network
I_G	The peak intensity corresponding to the presence of graphite bonds in the structure in the Raman spectra
I_{SP}	The peak intensity associated with sp-hybridized carbon chains in the structure in the Raman spectra
ML	Machine learning
NCGA	NanoCarbon Genome Approach
SAW	Surface acoustic wave
Sp hybridization	Linear structure in molecules

References

1. Paras; Yadav, K.; Kumar, P.; Teja, D.R.; Chakraborty, S.; Chakraborty, M.; Mohapatra, S.S.; Sahoo, A.; Chou, M.M.C.; Liang, C.-T.; et al. A review on low-dimensional nanomaterials: nanofabrication, characterization and applications. *Nanomaterials* **2023**, *13*, 160. https://doi.org/10.3390/nano13010160.
2. Levchenko, I.; Baranov, O.; Riccardi, C., et al. Nanoengineered carbon-based interfaces for advanced energy and photonics applications: a recent progress and innovations. *Adv. Mater. Interfaces* **2023**, 10, 2201739. https://doi.org/10.1002/admi.202201739.
3. Casari, C.S.; Milani, A. Carbyne: from the elusive allotrope to stable carbon atom wires. *MRS Commun*. **2018**, 8(2), 207-219. https://doi.org/10.1557/mrc.2018.48.
4. Shi, L.; Rohringer, P.; Suenaga, K. et al. Confined linear carbon chains as a route to bulk carbyne. *Nature Mater*. **2016**, 15, 634–639. https://doi.org/10.1038/nmat4617.
5. Duley, W.W.; Hu, A. Polyynes and interstellar carbon nanoparticles. *ApJ.* **2009**, 698(1). https://doi.org/10.1088/0004-637X/698/1/808.
6. Hu, A.; Rybachuk, M.; Lu, Q.-B.; Duley, W.W. Direct synthesis of *sp*-bonded carbon chains on graphite surface by femtosecond laser irradiation. *Appl. Phys. Lett.* **2007**, 91(13) 131906. https://doi.org/10.1063/1.2793628.
7. Flood, P.; Babaev, V.; Khvostov, V.; Novikov, N.; Guseva, M. Carbon material with a highly ordered linear-chain structure. In Polyynes: Synthesis, Properties, and Applications; Cataldo, F., Ed.; Taylor & Francis Group: Didcot, UK, **2005**; pp. 219–252. http://dx.doi.org/10.1201/9781420027587.ch11.
8. Fang, J.; Xie, M.; He, X., et al. Machine learning accelerates the materials discovery. *Mater. Today Commun.* **2022**, 33, 104900. https://doi.org/10.1016/j.jmst.2021.07.045.
9. Zhou, T.; Song, Z.; Sundmacher, K. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. *Eng.* **2019**, 5(6), 1017-1026. https://doi.org/10.1016/j.eng.2019.02.011.
10. Lee, J.; Park, D.; Lee, M.; Lee, H.; Park, K.; Lee, I; Ryu, S. Machine learning-based inverse design methods considering data characteristics and design space size in materials design and manufacturing: a review. *Mater. Horiz.*, **2023**,10, 5436-5456. https://doi.org/10.1039/D3MH00039G.
11. Badini, S.; Regondi, S.; Pugliese, R. Unleashing the Power of Artificial Intelligence in Materials Design. *Materials* **2023**, 16, 5927. https://doi.org/10.3390/ma16175927ю
12. Wang, J.; Wang, Y.; Chen, Y. Inverse Design of Materials by Machine Learning. *Materials* **2022**, 15, 1811. https://doi.org/10.3390/ma15051811.
13. Vivanco-Benavides, L.E.; Martínez-González, C.L.; Mercado-Zúñiga, C.; Carlos Torres-Torres, C. Machine learning and materials informatics approaches in the analysis of physical properties of carbon nanotubes: A review. *Comput. Mater. Sci.* **2022**, 201, 110939. https://doi.org/10.1016/j.commatsci.2021.110939.
14. Kalidindi, S.R.; Buzzy, M.; Boyce, B.L., Dingreville, R. Digital Twins for Materials. *Front. Mater.* **2022**, 9, https://doi.org/10.3389/fmats.2022.818535.
15. DePablo, J.; Jones, B.; Kovacs, C.; Ozolins, V.; Ramirez, A. The materials genome initiative: the interplay of experiment, theory, and computation. *Curr. Opin. Solid State Mater. Sci.* **2014**, 18(2), 99-117. https://doi.org/10.1016/j.cossms.2014.02.003.
16. Xie, J.; Su, Y.; Zhang, D.; Feng, Q. A vision of materials genome engineering in China, *Eng. J.* **2022,** 10, 10-12, https://doi.org/10.1016/j.eng.2021.12.008.
17. Qian, C.; Siler, T.; Ozin, G. Exploring the possibilities and limitations of a nanomaterials genome. *Small* **2015**, 11(1), 64-69. https://doi.org/10.1002/smll.201402197.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 February 2024 doi:10.20944/preprints202402.1117.v1

18. Ananyan, S.M.; Sazonov, D.S.; Slynko, Y.N.; Solomatin, E.B. Analytical Platform PolyAnalyst: Architecture, Functionality, Application Practice, 1st ed.; LLC "Hot Line - Telecom": Moscow, Russia, **2023**; 232 p., (in Russian). https://doi.org/10.25780/0008.
19. Ghiringhelli, L.M.; Vybiral, J.; Levchenko, S.V.; Draxl, C.; Scheffler, M. Big data of materials science: critical role of the descriptor, *Phys. Rev. Lett.* **2015**, 114, 105503. https://doi.org/10.1103/PhysRevLett.114.105503.
20. Yan, X.; Sedykh, A.; Wang, W., et al. Construction of a web-based nanomaterial database by big data curation and modeling-friendly nanostructure annotations. *Nat Commun.* **2020**, 11, 2519. https://doi.org/10.1038/s41467-020-16413-3.
21. Bornhoeft, L.R.; Castillo, A.C.; Smalley, P.R.; Kittrell, C.; James, D.K.; Brinson, B.E.; Rybolt, T.R.; Johnson, B.R.; Cherukuri, T.K.; Cherukuri, P. Teslaphoresis of carbon nanotubes. *ACS Nano* **2016**, 10, 4873–4881. https://doi.org/10.1021/acsnano.6b02313.
22. Hoglund, E.R.; Bao, D.L.; O'Hara, A., et al. Emergent interface vibrational structure of oxide superlattices. *Nature* **2022**, 601, 556-561. https://doi.org/10.1038/s41586-021-04238-z.
23. Streletskiy, O. A.; Zavidovskiy, I. A.; Nischak, O. Y.; Dvoryak, S. V. Electrical conductivity and structural properties of a-C:N films deposited by ion-assisted pulse-arc sputtering. *Thin Solid Films* **2020** 701, 137948. https://doi.org/10.1016/j.tsf.2020.137948.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.