Nominal digital twin for new generation product design

Haizhu Zhang[1], Rong Li[1], Guofu Ding[1], Shengfeng Qin[2], Qing Zheng[1*] and Xu He[1]

[1] *Institute of Advanced Design and Manufacturing, School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, 610031,P.R.China;*

[2] *School of Design, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK*

**Corresponding author. Email: qingzheng@swjtu.edu.cn*

Abstract： With the increasing competition of the manufacturing industry, it is essential for manufacturers to develop a new (or next) generation product based on their prior product design, product user experience, historical performances, etc. The digital twin (DT) is believed to be a suitable technology to support product lifecycle management with excellent data capture and analysis capabilities and product family design capabilities. However, it remains a challenge to synthesize and incorporate the data and information captured from previous generational products development and stored on their individual digital twin instances (DTIs) into next-generation product design. To address this problem, this study proposes a new nominal digital twin (NDT) concept as a collective digital representation of the current generational product digital mockups (DMUs) *and* all individual DTIs of the built products in services for new generational product development. NDT is first defined here as a prototypical and synthesized digital twin (or digital representation) of multiple individual digital twins corresponding to multiple physical products in use or previously used. By analyzing the data and information on their DTIs, NDT can enable the establishment and evolution of a more precise approximated model of many related family products used previously or currently in use in various application scenarios and environments in the physical world. This paper also demonstrates how a NDT model can be first established in the product design phase from various digital mockup models and enhanced later with a stochastic forest meta-model based on Bayesian optimization connected to DTIs. With this NDT model, collaborative exploration for optimal design solutions during new generation product design and improvement can be performed on NDT through multi-objective optimization, which in turn can make new generation product design easier and quicker. As a primary verification of the feasibility of our proposed approach, a case study has been carried out, and the results have well confirmed the our NDT-based new generation product design approach is feasible.

Key words： nominal digital twin; digital mock-up; closed-loop iterative; new generation product design; digital twin-based optimal design.

1 Introduction

In today's increasingly competitive markets, manufacturers are encouraged to use virtual product and manufacturing models, which are also called "digital mock-ups" (DMU)[1] and "digital twins" (DT)[2] in their entire product lifecycle. Without the need for costly and time-consuming physical mock-ups, these virtual models enable the efficient prediction of the consequences of product and procedure development as well as operation and maintenance choices based on the behavior and performance of current or past products[3-5]. Especially in design, such sophisticated virtual product models are crucial for early and efficient evaluation of the effects of design choices on the product's quality and functionality[6]. Current methods to implement DT models concentrate mostly on later product life stages such as production and operation[7-10]. However, it remains unclear how and in what ways communication, synergy, and coevolution between a physical object and its DMUs might result in a more informed, accelerated, and inventive design process [11, 12].

It has been argued that "a digital twin model serves as a set of linked operation data artifacts and simulation models, which is established after the as-manufactured physical entity" [13]. However, Grieves believes that the digital twin begins at the beginning of a new product's lifecycle and persists throughout the lifecycle[14]. In the create phase, the prototypical product with variants or all the products that can be built, is defined as a digital twin prototype (DTP). In figure 1, in the early stage of product development, the as-defined view linking between the product DMU and configuration management, provides the right 3D design data including 3D geometry, product structure and attributes for each product configuration or variant[15]; The as-designed is the successor of the as-defined in the development phase, functional DMU(FDMU) is proposed as a carrier, which can simulate what-if scenarios under various engineering domains and models, and hold the results of all simulations needed for full presentation of the product behavioral description including geometry, behavior and visualization of the simulated results[16]. Note that during the design stage, the as-defined and as-designed views may not have corresponding 3D physical models yet (sometimes, some 3D prototypes may be built for concept evaluation), and thus the as-defined and as-designed views cannot be regarded as digital twin models but only the DMUs. After each individual part or component is designed with manufacturing methods and specifications from its as-designed view and how all parts are assembled is defined, the resultant DMU is called Industrial Digital Mock-Up (iDMU) with more detailed manufacturing information for each part and the assembly. Up to this point, iDMU could be regarded as a DTP (prototypical digital Twin or digital Twin prototype) [14] or still as a digital model if *no intent* linking to the DTIs. Here we create a nominal digital twin (NDT) derived from iDMU and connected to DTIs and DTA (Digital Twin Aggregate). After an individual product is built, logistically transported and installed, it becomes a current physical product C-P_m, and its corresponding C-DTI_m will be built from the NDT with more detailed transportation, installation and location information just before being put into service, which continue traveling in a digital twin instance (DTI) journey during the product in use.

Therefore, when we look at the product design and development from a generational product point of view, many DTI models of the previous generation products have a large amount of data and information accumulated from the products can be twinning-back to update NDT by DTA. The updated NDT can be incorporated into the new generation product DMU (n-DMU) development in a close-loop fashion. Each DTI can provide 'real' working condition and performance data of a specific product in use. How to synthesize and incorporate this big data and information from these DTIs into new generation product design is a new research question that has not been well addressed before.

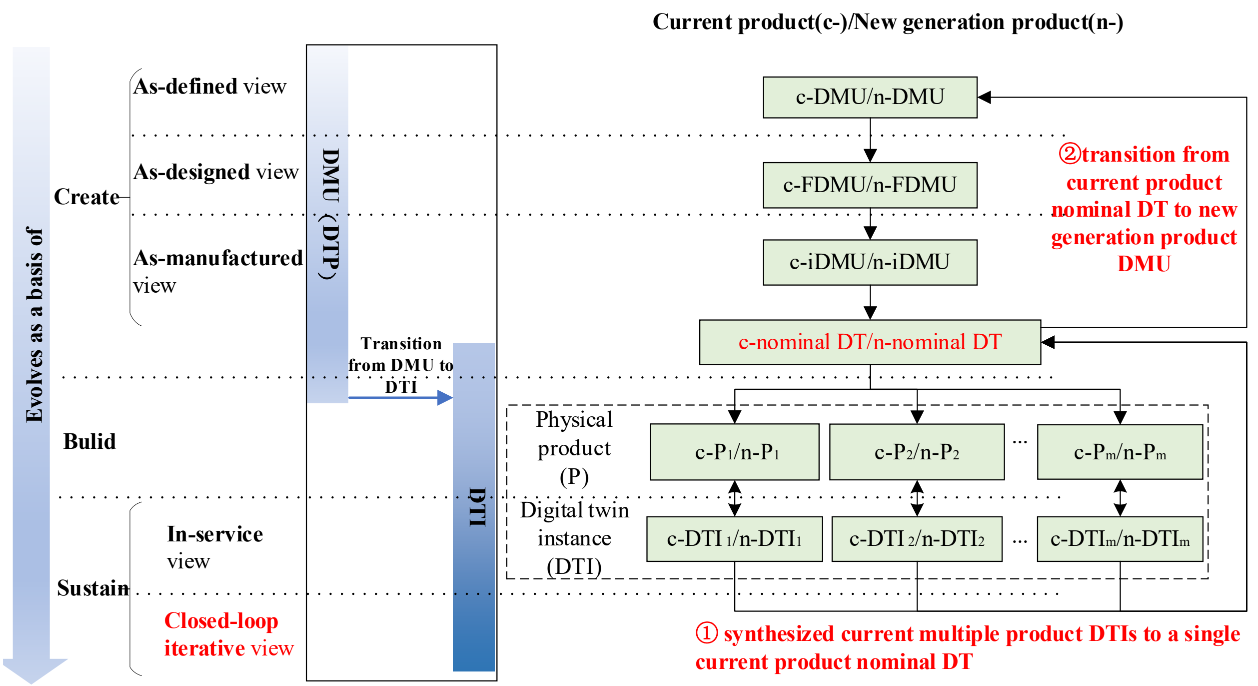

Figure 1 The closed-loop iterative design framework based on DMU and DT

Motivated by this need, this paper proposes a NDT-based method supporting a closed-loop generational product design. The loop starts from a single current product design's as-defined c-DMU, moves to its as-designed c-FDMU, and then moves to c-iDMU. When a physical product is produced and put into use, its individual c-DTI_i will be created and kept in updating status. Note that each c-DTIi could be different in terms of data captured from its product and running environment. Aggregating data and information on each c-DTI_i back into NDT for a close-loop generational product design is important for supporting product lifecycle innovation. Our solution is to synthesize current multiple product c-DTI_i into a single current product nominal DT (c-nominal DT) by linking a parametric (Meta) model driven by the real-world data on each c-DTI_i to an optimized approximate model in NDT. The main contributions of this paper are:

(1) We introduce a new concept of nominal digital twin (NDT), which can support (a) collaboration and cross-learning among individual instantiated product digital twins in the design-manufacturing-operation-maintenance-design loop and (b) the digital twin (or data) driven new generation product design. NDT synthesizes the multi-instantiated product DTIs' information to support the closed-loop new generation product design.

(2) We propose an NDT-based new generation product design method based our DTIs synthesis approach. Our approach is to link a parametric (Meta) model related to design research needs to some analysis of the real-world data to optimize parameters in the corresponding approximate model in NDT, which can support a new generation product design based on machine learning.

The rest of this paper is structured as follows: In the next section, DMU in product development and digital twin-driven product design are reviewed. After that, the proposed NDT-based new generation product design method is described. Its application in the development of a high-speed train's bogie is presented in section 4. Section 5 discusses the relationship between Nominal Digital Twin and Digital Twin Aggregate in our approach**.** Finally, conclusions are drawn.

2 Related work

2.1 DMU in the product development

The DMU has been defined in broader literature as a comprehensive digital representation of a product, component or system throughout the product lifecycle to reflect its geometry, function, performance, etc.[22, 23], which has the three technical characteristics of authenticity, full lifecycle view, and interdisciplinary. The basic purpose of a DMU is to reduce design and development time and promote inexpensive virtual prototyping[24]. DMUs are used by most firms in an effort to achieve design solutions as early as feasible in the product life cycle[1].

DMU expands the notion of digital prototyping beyond the design stage to production, maintenance, and recycling[1]. As defined by the European AIT, DMU is a computer simulation model based on a geometric product structure with full structural integrity[1, 25], which is a powerful verification tool for supporting including product functionality, assembly and maintenance process design, visualization, performance simulation, etc. throughout the entire lifecycle[26]. The as-defined view linking between the product DMU and configuration management is the configured DMU (CDMU), which provides the right 3D design data including 3D geometry, product structure and attributes for each product configuration or variant, and it has a single product structure for all the products within each configuration[15]. The as-designed and as-manufactured views are the successors of the as-defined in the series development phases. Many studies have been conducted to enhance DMUs for downstream processes, simulations, and system-level manufacturing support [27, 28]. In order to enrich the functions and behaviors of DMUs used in the early stage of digital product creation, a functional DMU (FDMU) is proposed as a carrier, which comprises the results of all simulations within various engineering domains[16]. Furthermore, Mas et al.[29, 30] present the Industrial Digital Mock-Up (iDMU) as a shared platform accessible to all product lifecycle stakeholders in order to achieve optimum design, and to address the difference between 'as-designed' nominally sized components and 'as-manufactured' actual components[17]. The as-manufactured iDMU is the complete digital definition of manufactured products[19], which can be transformed as instantiated iDMU for a single product and is iteratively formed by its part/component iDMUs. It can hold many actual data including product design, process design and manufacturing resource data[20]. Friel et al.[28] create an enriched DMU (EDMU) to be used to enable a designer to perform tolerance analysis in the CAD domain, which is a reflection of 'as manufactured' part forms allowing assembly analysis and tolerance consideration to be more accurate earlier in the design process.

In addition to geometric information, DMUs might serve as information models for other aspects of data. Kiritsis et al. [31] concentrate primarily on the multi-physics models of the product DMU in order to design and anticipate the structural life of the product and to improve methodologies for product certification and maintenance. Mas et al. [18, 30] concentrate on manufacturing models to solve difficulties associated with the integration of functional and industrial design for supporting a collaborative delivery to the producing, maintaining, and servicing of the product. For the purpose of model-based systems engineering (MBSE) deployment in the context of small and medium-sized businesses (SMEs), Chapurlat and Nastov [24] defined a formalization of a DMU and proposed a federating multi-viewpoint modeling approach to construct DMU(s) that can be processed in various ways to meet the requirements of various stakeholders.

Products with a high market value, such as high-tech machine tools, trains, wind turbines, etc., are often technologically complex, costly, and reliable, necessitating constant maintenance throughout their life cycles[4]. This necessitates improving product design and manufacturing utilizing in-service feedback[32] and stresses improved product design through the alignment of designers, producers, consumers, and recyclers. Maintenance, use, and decommissioning feedback data are crucial to data-driven design and simulation modeling for future scenarios at the BOL (Beginning of Life) phase of product. This demands a closed-loop generational product design methodology with lifecycle data for current products [31, 33, 34]. The progressively expanded product related data and information should be reflected in the DMU of the new product [35].

However, from the standpoint of connotation, DMU concentrates primarily on the representation of the geometry, function, and performance of a designed and built product, and does not include any references to the operation and maintenance phases. It is difficult for DMUs to enable a closed-loop, iterative, and data-driven product design paradigm in the absence of such links.

2.2 Digital twin-driven product design

DT is an integrated multi-physics, multi-scale, and probabilistic simulation of an as-manufactured product, it couples a physical product to its digital representation or digital shadow or twin [36]. If only one way communication exists from a physical product to its virtual counterpart, it is referred to as a digital shadow[37], and when the two ways communications are established automatically, it is a digital twin. Unlike DMUs that just concentrate on the virtual world, DT is defined by the bidirectional interactions between the virtual and real worlds. Digital twin models are built on DMUs, but they are greatly enhanced by their connections to the physical world, which allows for data accumulation and simulation model improvement along the product lifecycle[38]. The holistic use of digital twin models in product development will dominate future product development [39].

A new paradigm for data-driven product design has just developed[6]. Many frameworks for digital twin-based product design have been developed in the same vein; however, they all describe product design based on a single product digital twin. The stages of the product design process include conceptual design, detailed design, and virtual verification. After the three phases, the prototype will be obtained. Digital twin technology will be used throughout the whole procedure[12]. An intelligent vehicle's digital twin was conceptually modeled [40], under a proposed five-dimensional digital twin and it can capture both the product generated data and its customers generated data. Digital twins can also capture product running environmental factors[41]. Digital twin-driven virtual verification can help discover design defects and make quick modifications, and then improve the design scheme and cooperation and avoid lengthy verification and testing [42]. A general DT architectural reference model was presented in [43] to facilitate the efficient optimization of product families. To examine how digital twin technologies contribute to the design of smart manufacturing systems, Researchers [9] provide a novel function-structure-behavior-control-intelligence-performance framework. By using supervised learning to create a more accurate approximation of the physical world, as described in [44] with the concept of an evolutionary digital twin (EDT), a new method for intelligent industrial product creation is proposed.

From the above overview, existing studies on digital twin-driven product design are mainly focused on the framework. These frameworks did not pay attention to how to synthesize a potentially huge number of individual digital twins to support new generation product design. This is a more complex research issue.

2.3 Research gap and motivation

The main issue to be solved in new generation product design is to find a proper way to take various feedback information from many existing product DTIs in their later life stages into the early stage DMUs of the new generation product design, making use of all-round data and information from various DTs from product design to the manufacturing, operation and maintenance phases. This requires creating a nominal digital twin model to connect data and information embedded in DMUs in the create phase, with synthesized approximate models of products from individual physical product digital twins to support new generation product design. To sum up, existing studies on DT-based product design are limited to:

(1) The digital twin theory emphasizes the description of the real state of the product. Although the traditional product data management system can record, share and manage design drawings, models and documents, it only establishes a static and idealized product information model, which may vary with the actual state of each product. There are deviations in dynamic instance data such as machining, assembly, and inspection. Two urgent problems must be solved: how to establish a digital twin-based new generation product design model to more accurately describe and manage the real manufacturing and operation data of each instance product, and how to integrate it into a better approximated (or near ideal) product information model. (2) The core issue is the transition from the multi-DTIs of current products to a DMU of a new generation product design. The current product manufacturing, operation and maintenance data need to be dynamically fed back to the new generation product design, and an approximate information model that more accurately reflects product manufacturing, operation and maintenance status data needs to be established.

Here, we propose a new concept of nominal DT and a new generation product design method with it. This method not only facilitates the generation of DMUs for new generation product design but also helps meet the demand for product data information models.

3 NDT based new generation product design method

3.1 The definition of NDT and its features

In order to address the above issue, based on the typical definition of digital twin, a new broader definition is proposed to support the bidirectional transition and connection between DMUs and individual DTIs. The key features (see Fig. 1) of a NDT are twofold: (1) It has the same nominal product definition information created in the Design stage and embedded in DMUs; (2) It has nominal (approximate) data-driven application models such as simulation and prediction, which are synthesized/learned from multiple individual DTIs into a portfolio of models. The difference between an NDT and an individual DTI is that a DTI only has a single physical product related manufacturing, transportation, installation, work scenario, performance and behavior data, and its associated data driven application models are limited to the corresponding physical product.

Mainly based on machine intelligence and supplemented by human intelligence, the NDT could help effectively establish the behavior law under various uncertainties and scenarios and gradually achieve a well approximated model of the real-world products, namely the behavior module. When developing a new product, both design and design knowledge reuse are based on not only current product but also previous generation products in the same pedigree.

The application features of applying NDT to new generation product design, as shown in Fig.2, include some notable features:

(1) The NDT clearly builds an approximate digital world corresponding to the real world (where multiple products are in use in different environments, scenarios, and statuses).

(2) The NDT modeling is established by synthesizing different real world models with uncertainties in multiple virtual spaces in connections with the previous DMUs. The multi-DTIs are aggregated by establishing and synthesizing meta-models of multi-instanced products into a meta-model of NDT, which is detailed in Section 3.3.

(3) Compared with current digital twin which mainly uses a virtual space to predict the product behavior, the NDT can be constructed using an incremental development method that employs multiple DTIs for fast parallel learning and searching, which is helpful not only for fully understanding of the existing design solutions and but also for gaining insights for new generation product design. This is important to form a continuous product design and development platform that enables smart responses to market changes and sustains profit and growth.

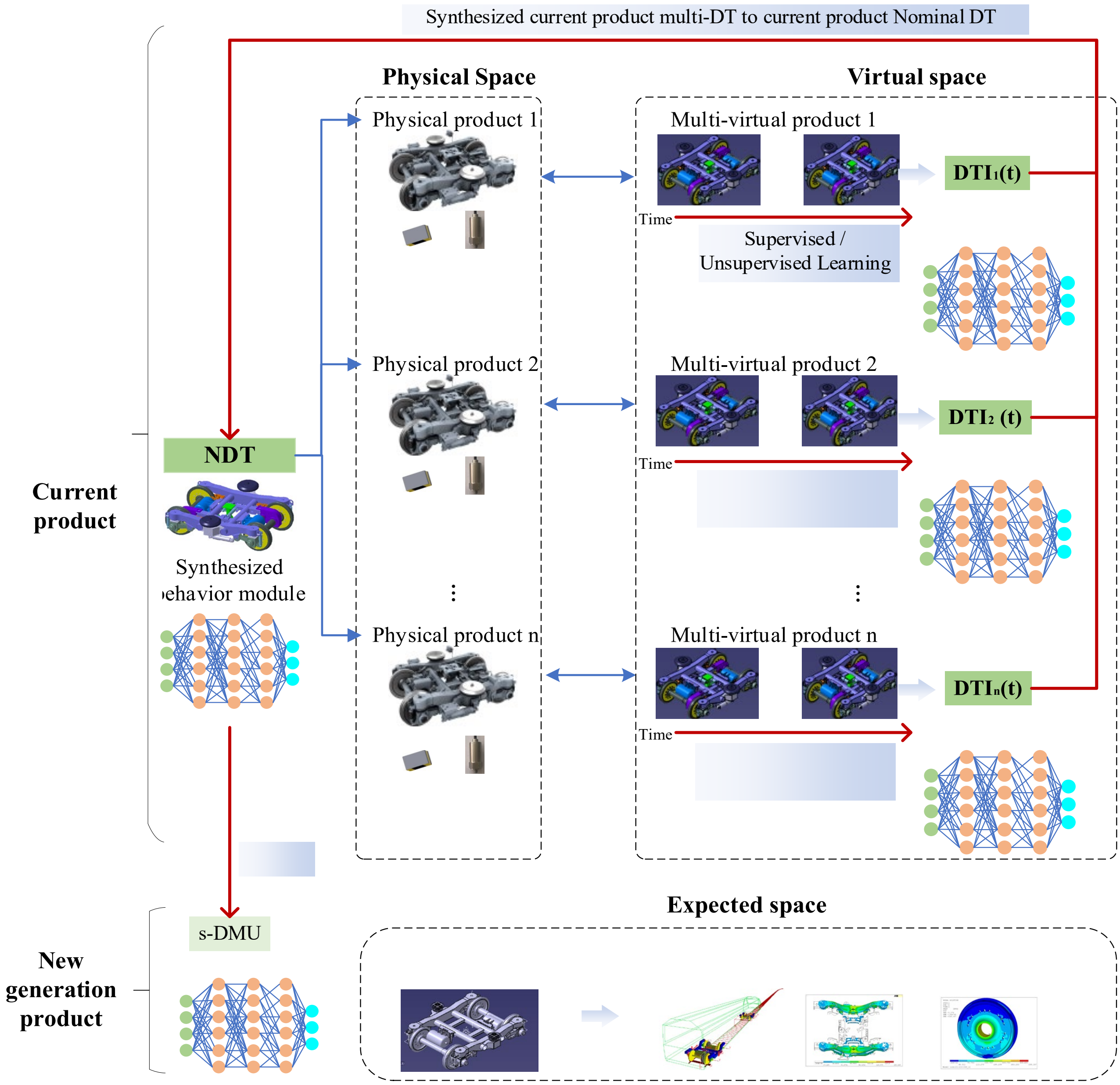

Figure 2 Core features of the NDT

3.2 Common closed-loop iterative product digital design framework

Enabled by cyber-physical systems, model-based system engineering, digital twin, and a growing endeavor for data gathering and processing, digital design is rapidly changing to form a closed-loop iterative product design framework as shown in Figure 3, leading to huge changes in the form and connotation of DMU. Prior to the digital design revolution, product design was predominantly a physical artifacts-oriented iterative design process. However, with the introduction of the CAE simulation, the way products were designed rapidly changed to digital mock-up model-oriented iterative design with virtual simulations and verifications. Recently, new design paradigms have been developed with the goal of combining design with manufacturing to achieve manufacture-oriented closed-loop iterative design. In order to achieve closed-loop new product iterative design (Fig 3) from design requirement and digital presentation(DP), the DMUs of the new generation product need to be established based on the nominal DT of current products, evolving from the multi-DTIs of current products to DMUs of the new generation product. Thus, the transition from multi-DTIs to NDT is the central problem in realizing the new generation product design in a closed-loop fashion. The detailed information for each DMU at different stages can be referred to the product lifecycle information model[45].

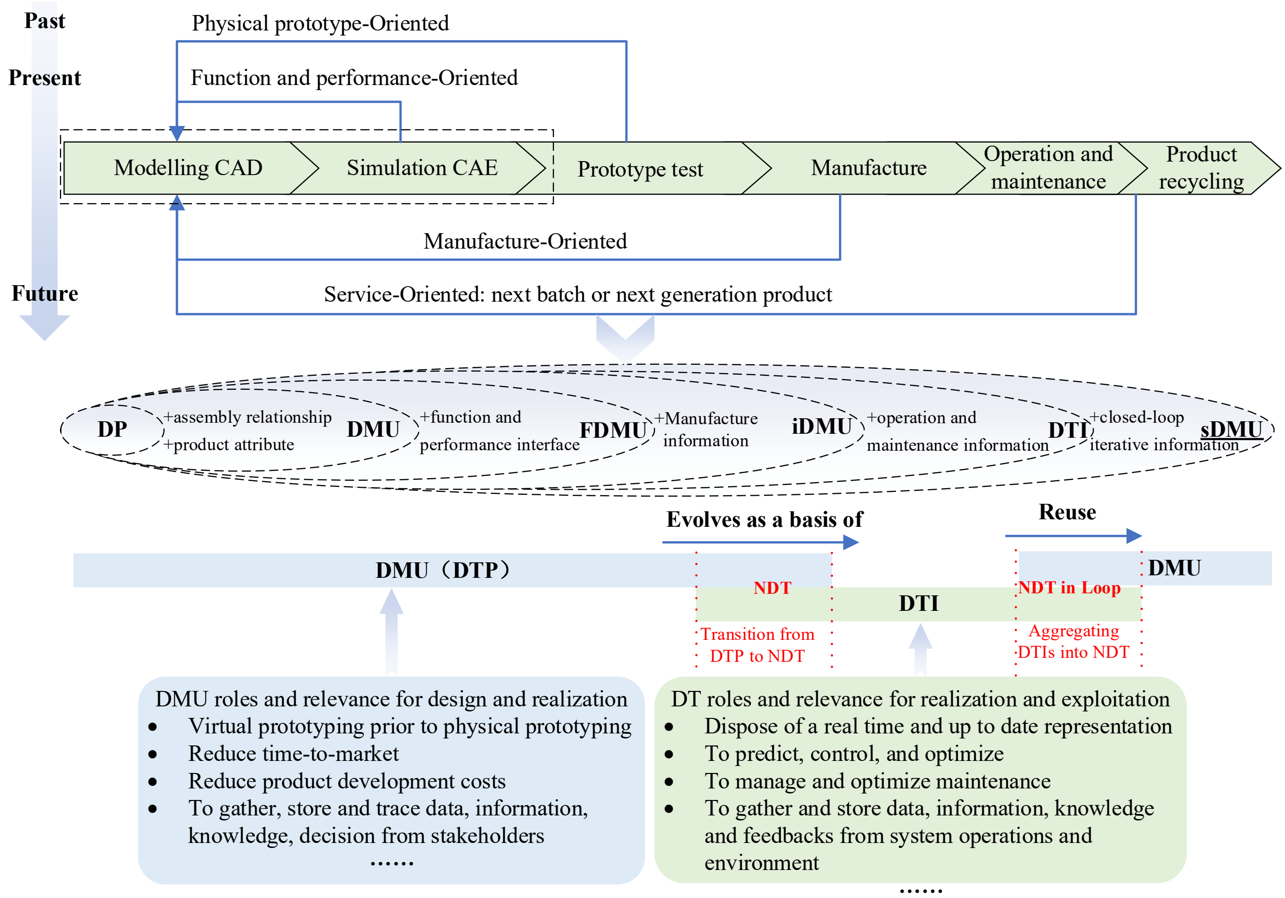

Figure 3 Trend of closed-loop digital product design

3.3 A new generation product design methodology

As explained before, the NDT acts as the engine that converts massive data handled by data lifecycle management into valuable information that can be directly used by designers to make wise decisions at various design stages, as required by design method and theory[11]. It is a depiction of both the designer's goals and the reality of the situation. According to the situated FBS framework[46], design involves back-and-forth interactions between three spaces: expected space, virtual/approximate space and physical space (Figure 4). In the primary design step, the expected space (corresponding to the ideal world in [44] and the expected world in [46]) is the world that is understood, imagined, and produced by designers. The essence of NDT is to create an approximate digital representation of the physical products in the virtual space and then reflect from the virtual space back to the expected space. This approximation model should be regarded as a dual-reflection of both the expected and physical spaces. The designers are directed to change their assumptions depending on the facts that are cross-examined in both the approximate and physical worlds, and to make better informed design choices as a result. Inconsistencies in the function, behavior, and structure of the expected product, the virtual product, and the physical product are gradually narrowed[11, 46].

How to establish this approximation model in NDT is a research problem. Here, we propose to link a parametric (Meta) model related to design research needs to some analysis of the real-world data to optimize parameters in the corresponding approximate model in NDT. Their relationships are shown in Fig 4.

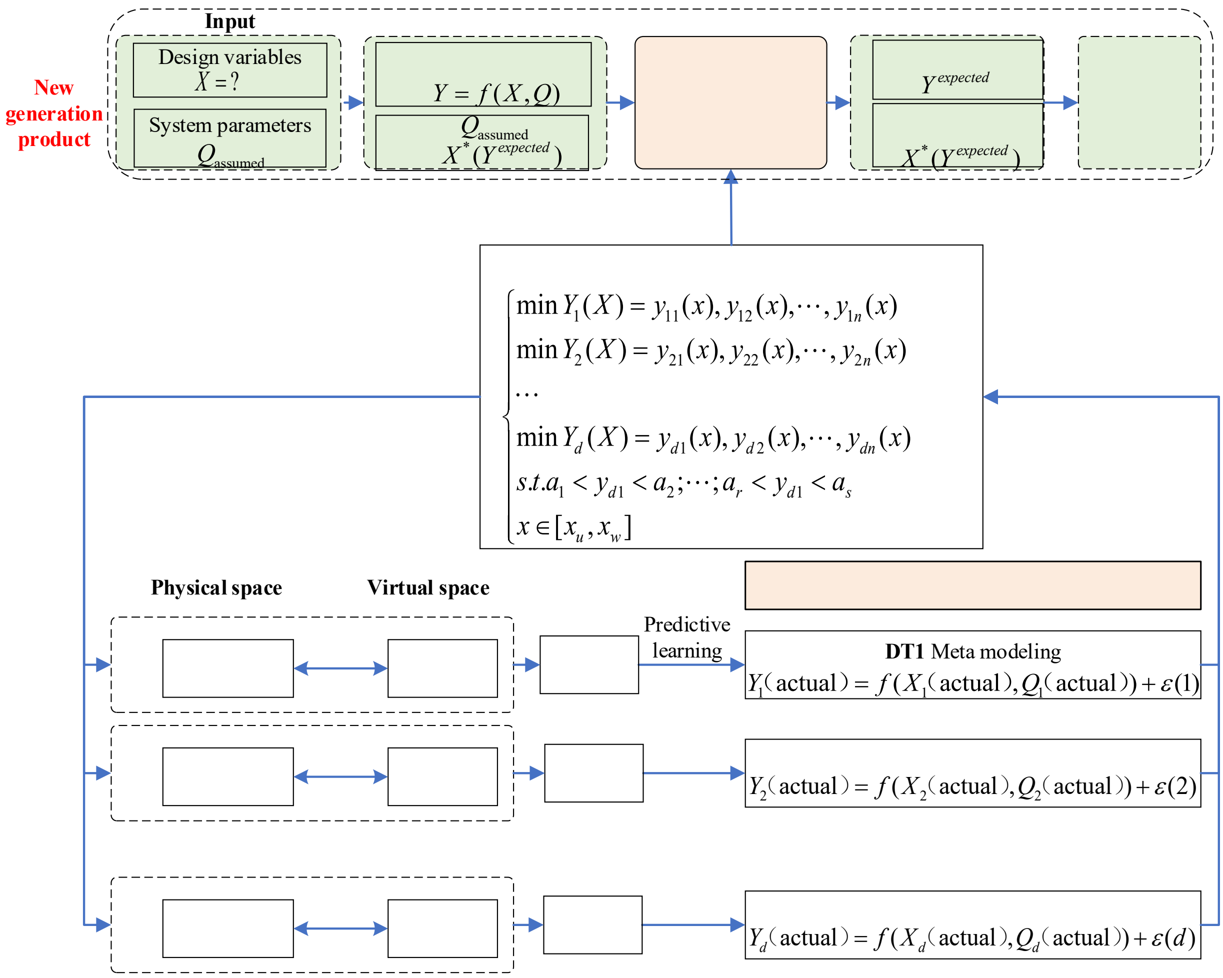

Figure 4 NDT-driven new generation product design framework

The aggregating the multi-DTIs to a nominal DT and then to a new generation product DMU (sDMU) is an inverse design problem. Considering the defects of DMUs in forward design, forward design and inverse design models interact iteratively through feedback[47, 48], based on which a combined method of inverse and forward design is proposed, as shown in Figure 4. This method helps to establish and synthesize meta-models of multi-instanced products into a meta-model of NDT, which aids in the development of new generation products through inverse optimization. The NDT contains a set of generalized/enhanced simulation models synthesized from multi-DTIs for the new generation product design, but other product definition information such as geometric information is the same as that defined in DMUs as a nominal product design.

In new generation product design, designers can use enhanced simulation models/capabilities in NDT based on a better understanding of relationships between the real physical and virtual worlds in new product design process, resulting in new as-defined, as-designed, and as-manufactured sDMUs for a new generation product definition.

The meta-model of individual digital twin instance (DTI) is established by stochastic forest meta-model based on Bayesian optimization. The multi-objective optimization mathematical model of multi-DTIs fusion is developed based on the constructed DTI meta-model. The aggregating method of multi-DTIs as follows: firstly, the optimization goal, optimization parameters, and optimization constraints of each instance condition are determined by analyzing the instance condition; Secondly, the ideal point achieved by single objective optimization is the outcome of single objective optimization that has been applied to each performance index of various instance situations. The relevance of the case conditions and the importance of the dynamic performance indices are considered while determining the correlation coefficients. Thirdly, the multi-objective optimization model is then modified based on the ideal point value and phase relation value, and the comprehensive multi-objective optimization model under multiple instance conditions is generated. The detailed method is descripted below.

（1）The forward design model of DMU generation

The original design problem is referred to as forward design, whose decision flow is shown in Fig.4. Assuming a forward design model:

$$Y = f(X, Q) \tag{1}$$

The system response $Y = \left(y_1, y_2, \cdots, y_l\right)$ is related to the product design specification at beginning of the design. There are certain system parameters including environmental factors as input conditions that are assumed to be known in priori, namely $Q=\left(q_1,q_2,\cdots,q_m\right)$, such as the expected capabilities of the product, the use scenarios of the product, the user's operating habits, etc. The optimal values of design variables $X=\left(x_1,x_2,\cdots,x_n\right)$ are X^* obtained based on the input of Q. The corresponding output of design is Y^{expected} with respect to the optimal design X^*. Because of the subjectivity of traditional trial evaluation such as load spectrum data acquisition based on experimental or prototype testing, it is difficult to know the actual product operating conditions prior to design optimization.

（2）NDT construction method

Each DTI serves to capture various information with uncertainties associated with a single product and simulate them in the digital world. Uncertainties are caused by incomplete and/or unknown information. Any product's actual performance in the physical world can be affected by a few factors with uncertainties, which could arise at different phases throughout a product's lifecycle (e.g., production, distribution, usage, maintenance, recycle, etc.). In particular, the aggregation of many factorial information with uncertainties may significantly affect a product's function, behavior, and structure. While NDT serves to synthesize DTIs captured data and information with various uncertainties in the physical world and simulate them in the digital world, so that more robust design solutions can be generated and virtually validated against the information with uncertainties. By comparing the virtual context and physical context, designers can deepen their understandings of the ideal and real contexts in which a product is used, and such understandings are especially important for designers to improve a product's adaptability. Considering the uncertainties influenced by manufacturing errors and other factors, i.e., ε, a product design model forms d physical product instances, the as-built product behavior models are as follows:

$$\begin{cases} Y_1 = f(X_1,Q_1)+\varepsilon(1) \\ Y_2 = f(X_2,Q_2)+\varepsilon(2) \\ \ldots\ldots \\ Y_d = f(X_d,Q_d)+\varepsilon(d) \end{cases} \tag{2}$$

In general, the NDT describes $Y^{\varepsilon}=f\left(X^{\psi},Q^{\omega}\right)$ the mapping relationship between X, Q and Y with uncertainties ψ, ω and ε respectively. The uncertainty profiles of ψ, ω and ε can be better estimated from equation (2). A meta-model (or surrogate model) that represents a family of tasks, which are used to simulate the behaviour of structures and provide an approximation of the original model's response in a fraction of time[49, 50]. This relationship can be described by a meta-model based on machine learning, such as a random forest model detailed below.

While following the forward design process, inverse design considers feedbacks of the product in use and takes input of identified new design needs. In other words, inverse design imposes a feedback loop to forward design. The product operational data mining produces useful product usage information regarding the actual system responses Y_{actual} corresponding to the product design based on, i.e., $X|Y_{\text{actual}}$, along with the actual observation of system parameters, i.e., Q_{actual}. It reflects the inverse relationship from Y_{actual} of system response and X^* of design parameters to infer what an optimal system parameter setting Q^* is supposed to be. Applying this design knowledge to improve design for new generation product development, $Q^*|\ Y_{\text{actual}}$ is brought back to the forward design process to replace the original assumption, i.e., $Q^*\rightarrow Q_{\text{assumed}}$, thus forming a closed loop of design decision to yield an improve design that is dedicated to the new needs of customers.

Current approaches to the implement of multi-DTI fusion lack of a model, which hinders the transition from the multi-DTIs to a NDT, i.e., DTI(n)→NDT (1).

The actual design parameters X subject to changes when a great number of products are in use and each individual product is different with constantly changing factors such as wearing, repairing and brand new or second-hand part replacement. The actual design parameters of individual product, which is expressed into vectors for multi-DTIs. The actual system parameters Q of product are different for individual product under different operation conditions and changing physical environments. The actual system parameters of individual product, i.e., Q_d, are expressed into a vector for multi-DTI. Based on the vector, the corresponding output of design is Y with respect to the optimal design X, its meta-model is established as follows:

$$\begin{cases} DTI_1: Y_1(\text{actual}) = f(X_1(\text{actual}), Q_1(\text{actual})) + \varepsilon(1) \\ DTI_2: Y_2(\text{actual}) = f(X_2(\text{actual}), Q_2(\text{actual})) + \varepsilon(2) \\ \ldots\ldots \\ DTI_d: Y_d(\text{actual}) = f(X_d(\text{actual}), Q_d(\text{actual})) + \varepsilon(d) \end{cases} \tag{3}$$

This paper intends to use random forest model[51] to solve the problem of meta-model construction for NDT. As the random forest model has the characteristics of easy implementation, low computing cost and good scalability, the random forest model can minimize the loss caused by data loss in the case of incomplete data, especially when the data loss is large, the random forest model can still maintain high fitting ability. When random forest is used for prediction, data sampling is firstly carried out to obtain the training set for establishing each decision tree. Then, a decision tree is constructed based on CART node splitting algorithm, and multiple decision trees constitute a random forest model. Finally, the average value of the predicted value of all decision trees is the predicted value of the random forest model. For the regression problem, the mean value of the results of k decision trees is calculated as the final result, and the expression is in Formula (4).

$$R(x) = \frac{1}{k}\sum_{i=1}^{k} T_i(x) \tag{4}$$

Where, k represents the number of decision trees in the random forest; $T_i(x)$ represents the result of the i_{th} decision tree in the random forest.

In this paper, a random forest meta model based on Bayesian optimization hyperparameter building method is proposed. In this method, various parameters in the original random forest are optimized to realize the optimization of the model to achieve the optimal model under different parameter combinations. By optimizing the parameters of the random forest regression function, the optimal model with different parameter combinations is obtained, and the prediction accuracy of the model is improved. The program uses the random forest model in the Python SciKit-learn module to set parameters and adjust the combinations continuously. The accuracy of the prediction model is obtained in the experiment, and the optimal parameters are obtained through Bayesian optimization, so as to find the optimal parameter combination of the random forest prediction model. Gaussian process here is the combination process of different parameters in the random forest model, and the linear combination of any finite number of samples can be expressed as a joint Gaussian distribution:

$$f(x) \sim gp(m(x), w(x, x')) \tag{5}$$

Where, $m(x) = E(f(x))$ is the $f(x)$ mathematical expectation, (x, x') is the x covariance function.

The mean and variance of $f(x)$ can be obtained by inputting data into the Gaussian model, and the Gaussian distribution of the function can be constructed. By increasing the amount of data, the gap between the predicted distribution and the real distribution can be narrowed.

The process of Bayesian optimization hyperparameter building method for random forest meta-model [52] is as follows:

Step1: Initialization parameters are randomly generated within the range of the number of random forest hyperparameters, and these initialization parameters are input into the Gaussian model. Test samples are then input into the fitting model to obtain model output, which is then modified to bring the model closer to the true distribution of the function.

Step2: For the modified Gaussian model to approach the real distribution of the objective function faster and more accurately than other combinations of candidate sets, the extraction function is used to extract the parameter combination points that need to be evaluated in the following step from the modified Gaussian model.

Step3: The algorithm ends and exits, and outputs the appropriate parameter combination and the model's prediction error $(x_i, f(x_i))$ when the error of the parameter combination satisfies the target requirements.

Step4: If $f(x_i)$ doesn't satisfy the requirements, add $(x_i, f(x_i))$ to the Gaussian model to change it, then repeat steps 2 and 3 until the predetermined accuracy requirements are satisfied. To assess the correctness of the meta-model, often Mean Squared Error (MSE), Mean Absolute Error (MAE), and determination coefficient (R-squared, R2) are utilized, as indicated in formula (6)- (8).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{6}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \hat{y}_i)| \tag{7}$$

$$R^2 = 1 - \frac{\sum_i(\hat{y}_i - y_i)^2}{\sum_i(\bar{y}_i - y_i)^2} \tag{8}$$

Where: $y_i - \hat{y}_i$ is the real value of the test set minus the predicted value.

(3) The inverse optimization design model with NDT

In order to build the multi-objective optimization synthesis model from a multi-instance fusion, firstly, the optimization goal, optimization parameters, and optimization constraints of each instance condition are determined by analyzing the instance condition; Secondly, the ideal point achieved by single objective optimization is the outcome of single objective optimization that was applied to each performance index of various instance situations. The relevance of the case conditions and the importance of the dynamic performance indices are considered while determining the correlation coefficients. Thirdly, the multi-objective optimization model is then modified based on the ideal point value and phase relation value, and the comprehensive multi-objective optimization model under multiple instance conditions is generated. Finally, in order to create an optimized non-inferior solution set, the optimization model is first translated into Python, after which a multi-objective intelligent optimization is carried out using the suitable optimization algorithm.

The multi-objective optimization mathematical model of a multi-instance fusion is developed as indicated in Formula (9) based on the previously constructed NDT meta-models. The mathematical representation of a solvable multi-objective optimization is created by transforming the multi-objective optimization model of multi-instances using the ideal point method and multi-objective programming method.

$$\begin{cases} \max/\min Y_1(X) = y_{11}(x), y_{12}(x), \cdots, y_{1n}(x) \\ \max/\min Y_2(X) = y_{21}(x), y_{22}(x), \cdots, y_{2n}(x) \\ \cdots \\ \max/\min Y_d(X) = y_{d1}(x), y_{d2}(x), \cdots, y_{dn}(x) \\ s.t. a_1 < y_{d1} < a_2; \cdots; a_r < y_{dn} < a_s \\ x \in [x_u, x_w] \end{cases} \tag{9}$$

Where, x is the design variable, which has a certain design range, and the lower limit of its parameter value is x_u and the upper limit is x_w ; $\max/\min Y_1(X)$, $\max/\min Y_2(X)$,…., $\max/\min Y_d(X)$ represent the optimization problem of different instances of working conditions; $y_{dn}(x)$ represents the value of each performance indicator in a specific instance.

Ideal point method is a kind of evaluation function method for solving multi-objective programming problems, mainly by making the target value as close as possible to the ideal value to solve, so as to obtain effective solutions. The ideal point determined by the principle of the ideal point method represents the ideal optimal solution of a single target obtained separately under the same constraints. We need to transform the performance indexes $y_{11}(x)$, y_{21} (x),⋯ y_{d1} (x) under different instance working conditions into single objective problems by ideal point method. The y_{11}^*, y_{21}^*, ⋯, y_{d1}^* represent the ideal optimal solution of a single target obtained separately under the same constraints. The optimization mathematical solution of many instances conditions is converted into the multi-objective optimization mathematical problem based on the formula (9) as follows:

$$\begin{cases} \min/\max y_1(X) = \sum_{d=1}^{D} \omega_d \dfrac{y_{d1}(X)}{y_{d1}^*} \\ \min/\max y_2(X) = \sum_{d=1}^{D} \omega_d \dfrac{y_{d2}(X)}{y_{d2}^*} \\ \cdots \\ \min/\max y_n(X) = \sum_{d=1}^{D} \omega_d \dfrac{y_{dn}(X)}{y_{dn}^*} \\ s.t. a_1 < y_{d1} < a_2; \cdots; a_r < y_{dn} < a_s \\ X = \{x_1, x_2, \cdots, x_m\} \in [x_u, x_w] \end{cases} \tag{10}$$

Where, $y_1(X), y_2(X), \cdots, y_n(X)$ respectively represent output performance indicators; $y_{d1}(X), y_{d2}(X), \cdots, y_{dn}(X)$ respectively represent the performance index values under different instance conditions; ω_d is the weighted coefficient to express the importance of its objective function; $y_{d1}^*(X), y_{d2}^*(X), \cdots, y_{dn}^*(X)$ represent the ideal optimal solution of a single target obtained separately under the constraints of different instances of working conditions.

(4) The new generation product design parameters generation and verification

In expected design space of new generation product, the design corresponding output of ideal new generation product is Y_{expected} with respect to the optimal design $Q_{\text{actual}}|Y_{\text{expected}}$ and $X^*|Y_{\text{expected}}$, i.e., $Y_{\text{expected}}=f(X^*|Y_{\text{expected}}, Q_{\text{actual}}\ |Y_{\text{expected}})$, the design parameter of ideal new generation product.

The key design parameters $X^*|Y_{\text{expected}}$ of a bogie as input design data of the new generation product, the performance indicators Y_{expected} is the optimization objective, and range of performance metrics as constraints. According to the characteristics of the multi-objective optimization model of current product NDT, it is necessary to select the appropriate intelligent optimization algorithm for obtaining the design solution of new generation product. i.e. a set of the design parameters for satisfying multiple operating conditions.

The DMU modeling process then uses the design parameters as an input. The CAD and CAE software forms the foundation of the sDMU. The sDMU represents actual physical design circumstances and characteristics more accurately. Designers can develop engaging simulation scenarios to successfully apply simulation testing on prototypes and, to the best of their ability, forecast how the physical items will behave in use. This method can quickly make adjustments and precisely identify design flaws, improving the design scheme effectively while avoiding time-consuming testing and verification.

4 Case study-high speed train’s bogie

4.1 Case background

The bogie is the most important part of high-speed trains, as shown in Figure 5, which is coupled with multiple discipline domains such as mechanics, electricity, hydraulics, and control. It undertakes the tasks of carrying, guiding, damping, driving, braking, etc., thus, it is the fundamental part affecting the speed improvement of the train. As an illustrative example, the high-speed train’s bogie is a typical complex product system that has the characteristics of changeable operating environment and configurable requirements. The prevailing practice of high-speed train’ bogie design is to optimize design parameters by means of performance simulation. The DMU modelling method of a high-speed train’s bogie is traditionally based on a priori knowledge and experience about the product use cases. These methods aim at one standard design and have an assumed product use case scenario (e.g., considering a worst case of the condition) and fail to consider the various uncertainties associated with a product, which ignore the diversity of individual use scenarios under different operating environments. Enabled by the digital twin and a growing endeavor for data gathering and processing, this virtual product model is increasingly enriched with production and operation data. Therefore, it is of practical significance to empower high-speed train robust design by incorporating the previous data of product usage cases.

4.2 Construction of high-speed train’s bogie DT

The construction of a virtual product and a physical product for high-speed train’s bogie is described in Figure 5. As the environment is simply an object with track line.

Geometric data and dynamic characteristic data are first gathered through product description and geometric measurements in order to develop the virtual product. The virtual product model's geometric and dynamic modeling both make use of the obtained data. Due to WebGL's support for both 3D modeling and dynamic modeling[53], the modeling tool was created based on these two modeling processes. As a result, it could merge these two components to create a singular virtual product model. Additionally, the parameterized model is used in the dynamic modeling process. In this model, the dynamics behavior of a virtual product model could be described by the meta-model.

The physical product's construction is less difficult than that of the virtual one. As shown in Figure 5, the bogie is fitted with several sensors including acceleration, displacement, force, temperature, differential pressure, and velocity sensor. These sensors gather measurement data including force, temperature, speed, displacement and accelerated velocity , that could be used to advance the development of the virtual product model.

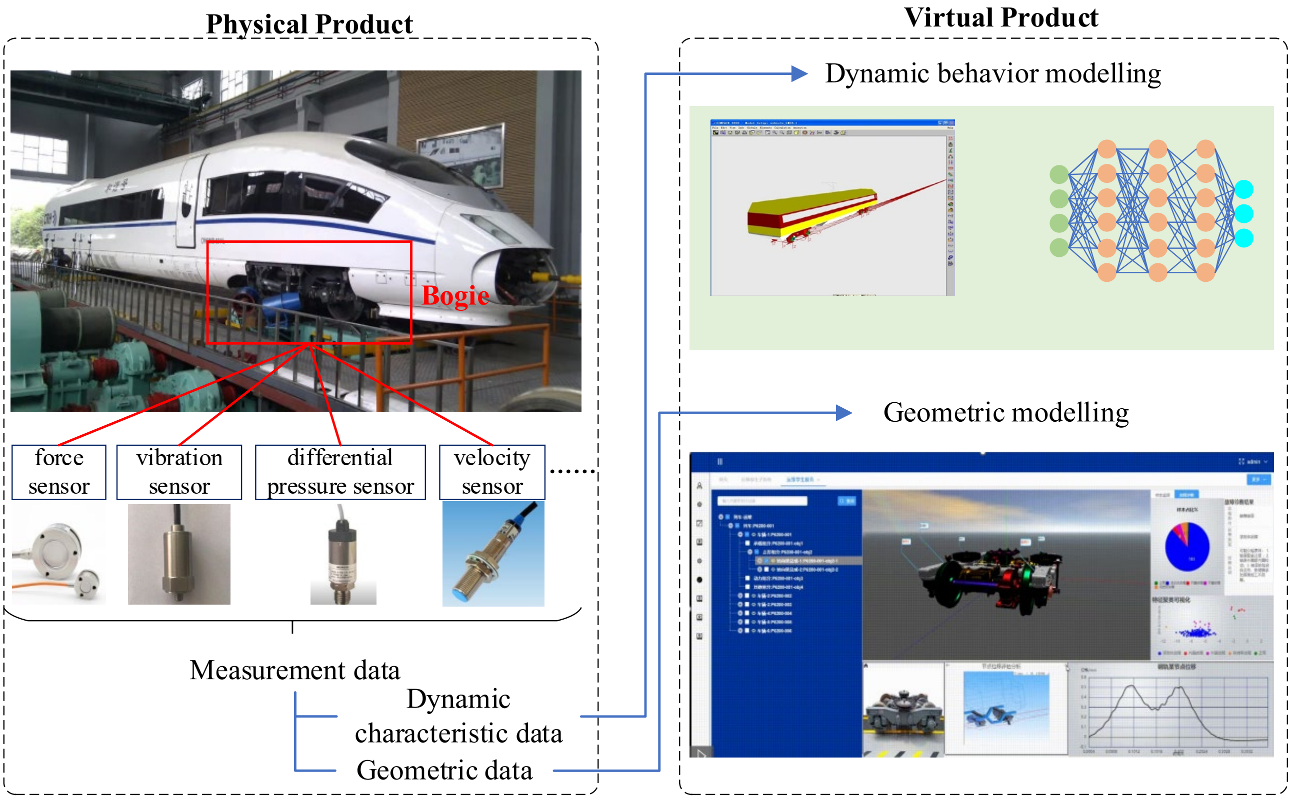

Figure 5 Physical product and virtual product model construction process

Along with the real operational environment for each market sector, the system parameters of a vehicle's dynamics are also different. The environment and line conditions are some of the factors that have the biggest effects on the performance of high-speed trains, such as rain, snow, ice, temperature, horizontal wind, and humidity. Environmental factors influence the model inputs for high-speed train dynamics, whilst the line condition has an impact on the stability and security of the dynamics. The actual design parameters of each train bogie are not the same when in use because of the influence of uncertainties in the machining process and the degradation of design parameters along its operation and maintenance journey. Some design parameters are also changing over time as a result of wear and fatigue. The dynamic performance of a high-speed train with different parameter combinations is different under different instance operating conditions. In the next generation product design, how to fast obtain the set of bogie design parameters corresponding to the dynamic performance of a high-speed train under multiple instances of working conditions based on the current product operation data is a critical issue, under the influence of uncertain factors, in order to reduce the complexity of optimization design and to improve design efficiency and design robustness.

4.3 The meta model of high-speed train's bogie NDT

Firstly, running twin data of 3 high-speed train's bogie instances, 15 key design parameters as the input X are identified by sensitivity analysis[54], 7 dynamic performance indicators response as the output Y, 4 service condition parameters as the system state parameter Q must be obtained. All of these parameters are listed in Table 1.

Table 1 Parameter sets of NDT meta-model

Parameter types	Parameter Name (Unit)	Range of parameters
Key design parameters of bogie(X)	x_1- wheel diameter（mm）	790-920
	x_2- Wheel set inner pitch（mm）	1350-1355
	x_3- Wheelset quality（kg）	1200-2200
	x_4- Rolling moment of inertia of wheelset（kg.m^2）	500-750
	x_5- Shaking moment inertia of Wheelset（kg.m^2）	500-800
	x_6- Longitudinal stiffness of primary spring（kN/m）	800-1150
	x_7-Vertical damping of primary suspension（kN.s/m）	10-30
	x_8- Longitudinal stiffness of axle box arm joints（MN/m）	5-15
	x_9- Lateral stiffness of axle box arm joints（MN/m）	4-10
	x_{10}- Lateral span of anti-snake shock absorbers（mm）	2400-2800
	x_{11}- Longitudinal stiffness of air spring（kN/m）	150-400
	x_{12}- Lateral stiffness of the air spring（kN/m）	150-400
	x_{13}- Vertical damping of secondary suspension（kN.s/m）	20-60
	x_{14}- Lateral damping of secondary suspension（kN.s/m）	30-50
	x_{15}- Series stiffness of anti-snake shock absorbers（MN/m）	5-13
Condition parameters of the instance(Q)	q_1- Running speed（km/h）	250-350
	q_2- Curve radius（m）	5000-9000
	q_3- super high（mm）	80-140
	q_4- track irregularities	Qin Shen spectrum, Beijing-Tianjin spectrum, Wu Guang spectrum
Dynamic response performance index(Y)	y_1- lateral stability≤	2.5
	y_2- vertical stability≤	2.5
	y_3- Axle vertical force≤（kN）	170
	y_4- Axle lateral force≤（kN）	55.7
	y_5- Derailment factor≤	0.8
	y_6- Wheel weight reduction rate≤	0.65
	y_7- overturning factor≤	0.8

(1) NDT meta-model construction

A high-speed train dynamic response design meta-model is constructed by using the Bayesian optimization random forest algorithm (BO-RF). Each instance is given 100 sets of data samples, and the first 90 sets of data in the data samples are subjected to BO-RF training. In this sample, the key design parameters of bogie (X) come from design data and actual measurement data of the manufactured product; the condition parameters of the instance (Q) come from sensor monitoring data and line measurement data; and the dynamic response performance index is from calculated results of sensor monitoring data and part of simulation analysis data. The values of the hyperparameters of the optimized random forest model are shown in Table 2. In this way, the dynamic design meta-model under multi-instance working conditions is obtained by training.

Table 2 Hyperparameter values of random forest model under different instance conditions

Random forest model	number of trees	Minimum number of samples required to split internal nodes	Minimum number of samples required for leaf nodes	maximum depth of tree
instance 1	132	3	15	9
instance 2	326	2	10	9
instance 3	271	4	7	12

(2) NDT metamodel accuracy verification

Use the other 10 sets of sample simulation data of each instance to verify the accuracy of the optimized hyperparameter model for the three instance working conditions. If the accuracy meets the requirements, save the meta-model. If it does not meet the requirements, add sample points and retrain the meta-model until the accuracy meets the requirements; the average absolute error, mean square error and coefficient determination of the random forest meta-model accuracy verification indicators are used to analyze and verify the fitting accuracy of the BO-RF algorithm of the instance. The results are shown in Table 3.

Table 3 BO-RF algorithm accuracy

Instance	Error	Lateral stability	Vertical stability	Axle vertical force	Axle lateral force	Derailment factor	Wheel weight reduction rate	Overturning factor	Comprehensive average
instance 1	MAE	0.1285	0.0083	0.5627	1.26207	0.0161	0.0157	0.0058	0.2856
	MSE	0.0253	0.0001	0.5190	2.4573	0.0005	0.0004	0.0001	0.4290
	R^2	0.9900	0.9979	0.9983	0.9949	0.9949	0.9986	0.9985	0.9962
instance 2	MAE	0.1189	0.0076	0.4867	0.9542	0.8755	0.0198	0.0056	0.3526
	MSE	0.0314	0.0001	0.6023	2.1462	0.0006	0.0004	0.0001	0.3973
	R^2	0.9895	0.9970	0.9984	0.9935	0.9854	0.9978	0.9982	0.9943
instance 3	MAE	0.1265	0.0104	0.6587	1.0243	0.8942	0.0201	0.0089	0.3919
	MSE	0.0412	0.0002	0.7564	2.6541	0.0008	0.0006	0.0001	0.4933
	R^2	0.9810	0.9899	0.9912	0.9886	0.9812	0.9823	0.9901	0.9863

According to the data of instance 1, the comparison results in the MAE, MSE and R^2 of the four common algorithms: back propagation neural network (BP), radial basis function neural network (RBF) and random forest (RF) and bayesian optimization random forest (BO-RF) are shown in Table 4.

Table 4 Accuracy comparison table of four algorithms

Algorithm	Error	Lateral stability	Vertical stability	Axle vertical force	Axle lateral force	Derailment factor	Wheel weight reduction rate	Overturning factor	Comprehensive average
BP	MAE	0.2052	0.0293	2.3044	3.7450	0.0406	0.0668	0.0255	0.9161
	MSE	0.0583	0.0013	6.9383	22.2772	0.0032	0.0062	0.0009	4.1757
	R^2	0.9770	0.9757	0.9775	0.9540	0.9688	0.9791	0.9818	0.9734
RBF	MAE	0.1519	0.0255	2.2714	2.2380	0.0358	0.0662	0.0217	0.6878
	MSE	0.0365	0.0010	6.8834	7.3647	0.0021	0.0059	0.0006	2.0499
	R^2	0.9856	0.9814	0.9773	0.9848	0.9793	0.9782	0.9746	0.9801
RF	MAE	0.1281	0.099	0.9419	1.5271	0.0234	0.0242	0.0083	0.3804
	MSE	0.0272	0.0002	1.4552	4.1593	0.0011	0.0010	0.0001	0.8063
	R^2	0.9892	0.9970	0.9952	0.9914	0.9886	0.9963	0.9967	0.9935
BO-RF	MAE	0.1285	0.0083	0.5627	1.26207	0.0161	0.0157	0.0058	0.2856

	MSE	0.0253	0.0001	0.5190	2.4573	0.0005	0.0004	0.0001	0.4290
	R^2	0.9900	0.9979	0.9983	0.9949	0.9949	0.9986	0.9985	0.9962

The above table shows that the BO-RF algorithm has high accuracy, the mean absolute error and mean square error are relatively small, the percentage of the error is less than 5%, and the fitting determination coefficient are all close to 1, indicating that the random forest meta-model based on Bayesian optimization hyperparameter has good accuracy, and the training achieves the effect. The meta-model is used to replace the coupled simulation model under the three instance conditions of the high-speed train, and the meta-model can be used for the multi-objective optimization design of new generation product.

4.4 NDT-driven new generation bogie design

The design of the new generation high-speed train's bogie should have higher robustness and adaptability, which is able to meet the needs of operation in different working conditions. In the NDT-driven new generation product design, the NDT model is established based on the twin data of multiple instances of the existing products. The optimal design parameters of the new generation product are solved by constructing a multi-objective optimization model from multi-instance fusions. In this model, initial 15 key design parameters of bogie as input design data for setting up the optimization design model, the 7 dynamic response performance indexes as optimization objectives, and a range of performance metrics as constraints. The obtained optimal design parameter solutions can make the new generation bogie meet the optimal dynamic performance indicators under the existing three case conditions (as shown in the table 5).

Table 5 Typical service conditions of the instance

Condition	Running speed（km/h）	Curve radius (m)	Super high(mm)	Track irregularities
Condition 1	250	5000	80	Qin Shen spectrum,
Condition 2	300	7000	100	Beijing-Tianjin spectrum
Condition 3	350	9000	140	Wu Guang spectrum

A multi-objective optimization model of three instance fusions is constructed according to formulas (9) and (10). Here we believe that the importance of the three instance conditions is the same, and the seven performance indicators of a single instance condition are all important indicators of safety and ride quality, which can also be regarded as equal in this optimization problem. If they are equally important, then their weight coefficients can all take the same value, and here the weight coefficient values can all take 1. The single-objective optimization is carried out by using the genetic algorithm, and the ideal value of each single-objective under the three working conditions is calculated separately. The specific data are shown in Table 6.

Table 6 ideal point value of three instance operating conditions

Ideal point value for case condition 1	y_{11}^{*}	y_{12}^{*}	y_{13}^{*}	y_{14}^{*}	y_{15}^{*}	y_{16}^{*}	y_{17}^{*}
	2.03	2.13	77.67kN	14.24 kN	0.12	0.28	0.27
Ideal point value for case condition 2	y_{21}^{*}	y_{22}^{*}	y_{23}^{*}	y_{24}^{*}	y_{25}^{*}	y_{26}^{*}	y_{27}^{*}
	1.98	2.03	75.56 kN	15.65 kN	0.13	0.27	0.29
Ideal point value for case condition 3	y_{31}^{*}	y_{32}^{*}	y_{33}^{*}	y_{34}^{*}	y_{35}^{*}	y_{36}^{*}	y_{37}^{*}
	2.12	2.19	79.54kN	16.34kN	0.18	0.29	0.30

The ideal point values in Table 6 are substituted into the optimization mathematical model established in formula (10), and the improved NSGA-II genetic algorithm is used for optimization calculation. Given a certain number of variables, the population size with the smallest convergence algebra is generally 4 to 6 times the number of design variables. Since the search convergence time of the algorithm is largely affected by the expansion of the population size, to obtain the optimal solution in a short time, we should not blindly select an excessively large population size. The initial population number in the multi-objective optimization in this paper is 80, the number of evolutionary iterations is set to 200, the crossover factor is 0.9, the cross-distribution index is 20, and the variation distribution index is 100. The Pareto optimal solution set obtained through multi-objective optimization is finally obtained. 80 sets of non-inferior solutions for the design parameters of high-speed trains are obtained.

To pick the ideal parameter set from the Pareto solution set as the parameters for the next generation of product design, designers or customers can select the required design parameter set from the solution set based on their preferences. In this study, the TOPSIS approach[55] is applied to quantitatively rank 80 non-dominated solutions, and the best 5 groups of solutions are chosen based on the ranking. The lateral and vertical stabilities are the second-level evaluation indicators of ride quality. The axle vertical force, axle lateral force, derailment factor, wheel weight reduction rate and overturning factor comprehensive average are the second-level evaluation indicators of safety. Safety is the basic requirement of high-speed train operation. Therefore, based on design experience and subjective preferences, the relative weights for ride quality and safety are set at 0.3 and 0.7, respectively. The decision matrix and normalization matrix are constructed based on 80 non-dominated solution set. Then, the positive and negative ideal solutions are constructed. Afterward, the positive and negative ideal solutions are constructed, and the Euclidean distance method is used to calculate the distance between each non-dominant solution and the positive and negative ideal solutions, as well as the relative closeness to the positive ideal solutions. The group of parameters with the maximum relative closeness is regarded as the best solution. From the 80 sets of non-inferior solutions, the selected 5 optimal parameter sets are shown in Table 7. The 15 key design parameters of DMU are the design result of new generation bogie.

Table 7 The optimal 5 groups of design parameter values for the 15 key design parameters

No.	Parameter name	Unit	Parameter symbol	1	2	3	4	5
1	wheel diameter	mm	x_1	803.32	815.43	825.37	855.47	860.36
2	Wheel set inner pitch	mm	x_2	1351.53	1352.17	1352.76	1354.64	1354.82
3	Wheelset quality	kg	x_3	1842.34	1876.56	1934.57	1980.42	2010.42
4	Rolling moment of inertia of wheelset	$kg \cdot m^2$	x_4	537.12	575.31	590.45	620.81	675.32
5	Shaking moment inertia of Wheelset	$kg \cdot m^2$	x_5	528.46	580.34	601.23	641.38	690.53
6	Longitudinal stiffness of primary spring	kN/m	x_6	984.97	827.54	969.73	982.37	931.40
7	Vertical damping of primary suspension	kN.s/m	x_7	15.32	18.43	20.56	13.74	21.58
8	Longitudinal stiffness of axle box arm joints	MN/m	x_8	6.83	7.24	7.95	9.47	10.24
9	Lateral stiffness of axle box arm joints	MN/m	x_9	4.51	6.43	7.54	8.55	7.82
10	Lateral span of anti-snake shock absorbers	mm	x_{10}	2521.34	2671.46	2643.21	2540.42	2489.65
11	Longitudinal stiffness of air spring	kN/m	x_{11}	193.34	183.99	202.04	194.47	192.81
12	Lateral stiffness of the air spring	kN/m	x_{12}	195.64	199.74	205.65	199.38	210.45
13	Vertical damping of secondary suspension	kN.s/m	x_{13}	26.43	33.65	28.46	30.66	36.84
14	Lateral damping of secondary suspension	kN.s/m	x_{14}	36.31	38.48	40.36	42.53	39.65
15	Series stiffness of anti-snake shock absorbers	MN/m	x_{15}	7.83	8.42	9.52	8.94	9.72

4.5 Results and discussions

The 5 sets of design parameters are substituted into the simulation model of the high-speed train system under 3 different working conditions for dynamic simulation calculation, and the corresponding 7 dynamic performance index values are obtained from the simulation. The values are compared with the dynamic performance index value corresponding to the initial design data of a certain type of CRH vehicle. The specific data comparison is shown in Table 8. In order to intuitively compare the dynamic performance of a certain type of CRH vehicle with the optimized design, there are 5 sets of solutions corresponding to the dynamic performance.

Table 8. Performance comparison table between the optimized solution and the original parameters of a certain type of CRH

Working conditions	Performance	A certain type of CRH	1	2	3	4	5

Instance 1	y_1	2.35	2.32	2.11	2.28	2.26	2.31
	y_2	2.07	1.98	2.03	2.06	1.99	2.10
	y_3	95.81	78.52	78.55	73.65	76.82	78.95
	y_4	25.15	14.15	12.88	13.54	13.95	16.53
	y_5	0.21	0.18	0.15	0.19	0.15	0.18
	y_6	0.32	0.19	0.19	0.22	0.26	0.28
	y_7	0.35	0.25	0.26	0.26	0.28	0.31
	Promotion	None	21.56%	24.84%	19.89%	19.91%	12.91%
Instance 2	y_1	2.20	2.18	2.22	2.21	2.19	2.12
	y_2	2.01	1.99	2.05	2.02	1.99	2.07
	y_3	78.30	75.43	79.65	80.23	74.62	72.65
	y_4	22.14	19.26	21.12	21.65	17.62	21.13
	y_5	0.18	0.17	0.19	0.18	0.18	0.21
	y_6	0.22	0.19	0.21	0.21	0.23	0.17
	y_7	0.26	0.24	0.22	0.23	0.28	0.25
	Promotion	None	6.49%	2.05%	2.13%	2.05%	3.19%
Instance 3	y_1	2.36	2.20	2.25	2.28	2.30	2.10
	y_2	2.21	2.02	2.06	2.07	2.11	1.99
	y_3	97.31	81.65	82.36	80.43	89.57	78.64
	y_4	29.98	24.31	25.65	24.84	26.34	20.34
	y_5	0.37	0.25	0.29	0.28	0.30	0.22
	y_6	0.41	0.29	0.30	0.29	0.33	0.24
	y_7	0.43	0.31	0.31	0.33	0.35	0.27
	Promotion	None	20.01%	16.80%	17.29%	12.03%	27.36%

According to the simulation analysis comparison of five optimization schemes under three different working conditions, the dynamic performance indicators of the five groups of design parameter sets not only meet the requirements of dynamic design standards, but also have been improved compared with the original design scheme. The new generation high-speed train's bogie has higher robustness and adaptability resulted from the NDT driven design method, which meets the needs of operation in different conditions. The result shows that NDT can enable the establishment of a more precise approximated model from the data and information not only related to a single product-in-use, but also to multiple physical products' performances and behaviors under a wide range of application scenarios in the physical world. Thus, it is proved that NDT-driven new generation product design method is feasible.

From the comparison results with other 3 algorithms in section 4.3, our Bayesian optimization random forest algorithm has high accuracy. Now, the new generation high-speed train's bogie is not built yet, thus, here testing method is based on the simulation analysis comparison of five optimization schemes under three different working conditions. Although there are some errors between the simulation results and the actual physical experiments, but the simulation results can prove that our method is feasible. Further physical verification is required in the future for evaluating its effectiveness.

5. Discussion on Nominal Digital Twin and Digital Twin Aggregate

As pointed by Michael Grieves in [14], the "twin" metaphor makes some believe that "digital twin exists ONLY after there is a physical product" although it is fallacy. In essence, Digital Twin model focused on the information about a product being populated and consumed from a logically centralized source across the four phases of a product's lifecycle: create, build, operate/sustain, and dispose phases. The DT does exist prior to a physical product and it may just has a different name such as digital model and the digital design [14]. Along a single product lifecycle from the Create, Build, Operate/Sustain to Dispose, Michael Grieves defined three types of Digital twin at the Macrolevel namely Digital Twin Prototype (DTP), Digital Twin Instance (DTI) and Digital Twin Aggregate (DTA) [14]. DTP is defined in Create (design) phase and used in all phases, DTI is defined in Build phase and used in the this and later phases, and DTA is defined and used in Operate/Sustain stage. The key differentiator of whether a digital model and associated information is a digital twin in Create phase is that it is intended that this model become a physical product and that its physical counter is realized" [14].

The paper [56] authored by Michael Grieves and John Vickers in 2017, only defines two types of digital twins in terms of DTP and DTI. DTA is defined in [14] referring to all the products that we did build. Its relationships with DTIs are clear, but its position in Sustain/service phase is not very clear. Based on our understanding, because each DTI in the service phase, the data goes both ways from a physical product to its virtual model or vice versa. So, each DTI can have a different data set related to its behaviour, performances, running environments and service scenarios. Aggregating all DTIs together can provide a big picture of the current product behavior and performance in general by synthesizing all DTIs. For example, we can aggregate information over a range of virtual and physical systems (twins) to correlate specific state changes with the high probability of future failures [56] through synthesizing the failure prediction model. But where this aggregating action could be embedded is not clear.

The DTP has all the information that will be needed to describe and produce a physical version that duplicates or twins the virtual version [56]. The information associated with a DTP usually includes but not limited to, the product requirements, CAD models (Fully annotated 3D models), Bill of Materials (BoM), behavioral simulations and any other information that would be needed to fully describe this new product. In addition, DTP would also include information needed to build this new product such as Bill of Process, Bill of Services, Bill of Manufacturing Systems, quality control information, and manufacturing simulations linking AI and M&S [14]. With reference to this criterion, not all digital models produced in the Create phase can be named as a digital twin. The relationships between digital twin and other digital models are not clear indicated.

Thus, in this paper, we look at the Create (Design) phase on a micro level and define the term "Nominal Digital Twin" (NDT) in the design phase prior to a physical product (see Fig 1). First, the NDT term indicates it is a DTP type of digital twin prior to a physical product by the term "nominal"; second, it shows a way of how to construct a NDT from digital Mock-Up models in sequence as shown in Fig 1, from As-defined, As-designed to As-manufactured prior to a physical product and intent to link to physical products late; third, it integrates aggregating all DTIs action into itself clearly demonstrating how to link individual product DTs (or DTIs) back to NDT by synthesizing DTIs for updates of NDT in right time and supporting next generation product design through the updates of early generational digital Mock-Up models. In this sense, NDT is a hybrid DT of DTP and DTA, but positioned as a DTP clearly.

6 Conclusions and future work

Motivated by the core idea of developing new generation product based on previous product DTs, we proposed an NDT approach, which is significant for acquiring the robust and optimal DMU of the new generation product. Based on the proposed NDT concept, a new closed-loop product design approach based on NDT is presented as a development framework to gain the optimal solution, and to guide the DMUs modelling.

Our case study on a high-speed train's bogie design shows that (1) the proposed NDT concept and associated design approach are feasible and applicable in new generation product design and (2) the proposed meta-model and optimization model are useful tools to realize the effective modelling and design.

Our future work will concentrate on the following two aspects:

(1) How to construct the NDT model to support design decision-making with uncertainties development in the closed-loop iterative design framework.

(2) How to use the NDT to guide individual instantiated DTs in manufacturing, use, and maintenance stages to learn from each other and have a collective learning capability to support application scenarios and conditions-based smart product performance.

Funding This work was supported by National Key Research and Development Program of China [grant number 2020YFB1708000]; Sichuan Science and Technology Support Project [grant number 2021YFG0039 and 2022YFG0252]; and National Natural Science Foundation of China [grant number 52105277].

Author contribution Haizhu Zhang contributed in the initial research idea and paper writing; Rong Li and Guofu Ding contributed to the conception of the study; Shengfeng Qin contributed in the nominal digital twin concept, the paper writing and proofreading; Qing Zheng and Xu He performed the experiment and performed the data analyses.

Declarations

Ethical approval and consent to participate Not applicable.
Consent for publication All the authors have given their consent for the publication of this manuscript.
Competing interests The authors declare no competing interests.

References

[1] B. Rooks(1998) A shorter product development time with digital mock-up. Assembly Autom 18:34-+. https://doi.org/10.1108/01445159810201405

[2] G. Michael(2014) Digital Twin: Manufacturing Excellence through Virtual Factory Replication. Whitepaper. https://doi.org/10.5281/zenodo.1493930

[3] M.C. Leu, H.A. ElMaraghy, A.Y.C. Nee, S.K. Ong, M. Lanzetta, M. Putz, W.J. Zhu, A. Bernard(2013) CAD model based virtual assembly simulation, planning and training. Cirp Ann-Manuf Techn 62:799-822. https://doi.org/10.1016/j.cirp.2013.05.005

[4] R. Roy, R. Stark, K. Tracht, S. Takata, M. Mori(2016) Continuous maintenance and the future - Foundations and technological challenges. Cirp Ann-Manuf Techn 65:667-688. https://doi.org/10.1016/j.cirp.2016.06.006

[5] C. Semeraro, M. Lezoche, H. Panetto, M. Dassisti(2021) Digital twin paradigm: A systematic literature review. Comput Ind 130. https://doi.org/10.1016/j.compind.2021.103469

[6] F. Tao, A. Liu, T. Hu, A.Y.C. Nee, D. Science(2020) Digital twin driven smart design. Academic Press, Amsterdam.

[7] F. Tao, Q.L. Qi(2019) Make more digital twins. Nature 573:490-491. https://doi.org/10.1038/d41586-019-02849-1

[8] Y.L. Wei, T.L. Hu, T.T. Zhou, Y.X. Ye, W.C. Luo(2021) Consistency retention method for CNC machine tool digital twin model. J Manuf Syst 58:313-322. https://doi.org/10.1016/j.jmsy.2020.06.002

[9] J.W. Leng, D.W. Wang, W.M. Shen, X.Y. Li, Q. Liu, X. Chen(2021) Digital twins-based smart manufacturing system design in Industry 4.0: A review. J Manuf Syst 60:119-137. https://doi.org/10.1016/j.jmsy.2021.05.011

[10] X.J. Niu, S.F. Qin(2021) Integrating crowd-/service-sourcing into digital twin for advanced manufacturing service innovation. Adv Eng Inform 50. https://doi.org/10.1016/j.aei.2021.101422

[11] F. Tao, F.Y. Sui, A. Liu, Q.L. Qi, M. Zhang, B.Y. Song, Z.R. Guo, S.C.Y. Lu, A.Y.C. Nee(2019) Digital twin-driven product design framework. Int J Prod Res 57:3935-3953. https://doi.org/10.1080/00207543.2018.1443229

[12] F. Tao, J.F. Cheng, Q.L. Qi, M. Zhang, H. Zhang, F.Y. Sui(2018) Digital twin-driven product design, manufacturing and service with big data. Int J Adv Manuf Tech 94:3563-3576. https://doi.org/10.1007/s00170-017-0233-1

[13] B. Schleich, N. Anwer, L. Mathieu, S. Wartzack(2017) Shaping the digital twin for design and production engineering. Cirp Ann-Manuf Techn 66:141-144. https://doi.org/10.1016/j.cirp.2017.04.040

[14] M. Grieves(2022) Intelligent digital twins and the development and management of complex systems [version 1; peer review: 4 approved]. Digital Twin 2. https://doi.org/10.12688/digitaltwin.17574.1

[15] R. Garbade, W.R. Dolezal(2007) DMU@Airbus - Evolution of the digital mock-up (DMU) at airbus to the centre of aircraft development. Future of Product Development:3-+. https://doi.org/10.1007/978-3-540-69820-3_2

[16] S. Fukuda, Z. Lulic, J. Stjepandic(2013) FDMU - Functional Spatial Experience beyond DMU? 20th Ispe International Conference on Concurrent Engineering:431-440. https://doi.org/10.3233/978-1-61499-302-5-431

[17] I. Friel, J. Butterfield, A. Marzano, T. Robinson(2017) Intelligent DMU creation: Toleranced part modelling to enhance the digital environment. Complex Systems Engineering and Development 60:92-97. https://doi.org/10.1016/j.procir.2017.01.038

[18] F. Mas, J.L. Menendez, M. Oliva, J. Rios, A. Gomez, V. Olmos(2014) iDMU as the Collaborative Engineering engine Research experiences in Airbus. Int Ice Conf Eng.

[19] J.L. Menendez, F. Mas, J. Servan, J. Rios(2012) Virtual Verification of an Aircraft Final Assembly Line Industrialization: An Industrial Case. Key Eng Mater 502:139-+. https://doi.org/10.4028/www.scientific.net/KEM.502.139

[20] J.L. Menendez, F. Mas, J. Servan, R. Arista, J. Rios(2013) Implementation of the iDMU for an aerostructure industrialization in AIRBUS. Procedia Engineer 63:327-335. https://doi.org/10.1016/j.proeng.2013.08.179

[21] M. Ito, M. Kamiya, A. Lujan(1984) Fluctuation of ELISA and skin biopsy results in individual inhabitants re-examined after several months in the endemic area of Guatemalan onchocerciasis. Ann Trop Med Parasitol 78:553-555. https://doi.org/10.1080/00034983.1984.11811864

[22] B. Johnston, T. Bulbul, Y. Beliveau, R. Wakefield(2016) An assessment of pictographic instructions derived from a virtual prototype to support construction assembly procedures. Automat Constr 64:36-53. https://doi.org/10.1016/j.autcon.2015.12.019

[23] Z.X. Li, X.P. Yan, C.Q. Yuan, Z.X. Peng, L. Li(2011) Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mech Syst Signal Pr 25:2589-2607. https://doi.org/10.1016/j.ymssp.2011.02.017

[24] V. Chapurlat, B. Nastov(2020) Deploying MBSE in SME context: revisiting and equipping Digital Mock-Up. 2020 6th Ieee International Symposium on Systems Engineering (Ieee Isse 2020).

[25] S. Aromaa, K. Vaananen(2016) Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design. Appl Ergon 56:11-18. https://doi.org/10.1016/j.apergo.2016.02.015

[26] P.G. Maropoulos, D. Ceglarek(2010) Design verification and validation in product lifecycle. Cirp Ann-Manuf Techn 59:740-759. https://doi.org/10.1016/j.cirp.2010.05.005

[27] A. Shahwan, J.C. Leon, G. Foucault, M. Trlin, O. Palombi(2013) Qualitative behavioral reasoning from components' interfaces to components' functions for DMU adaption to FE analyses. Comput Aided Design 45:383-394. https://doi.org/10.1016/j.cad.2012.10.021

[28] I. Friel, J. Butterfield, T.T. Robinson, A. Marzano(2020) Tolerance aware product development using an enriched hybrid digital mock up. Cirp J Manuf Sci Tec 30. https://doi.org/10.1016/j.cirpj.2020.04.009

[29] F. Mas, J.L. Menendez, M. Oliva, J. Rios(2013) Collaborative Engineering: an Airbus case study. Procedia Engineer 63:336-345. https://doi.org/10.1016/j.proeng.2013.08.180

[30] F. Mas, J. Rios, J.L. Menendez, A. Gomez(2013) A process-oriented approach to modeling the conceptual design of aircraft assembly lines. Int J Adv Manuf Tech 67:771-784. https://doi.org/10.1007/s00170-012-4521-5

[31] D. Kiritsis, A. Bufardi, P. Xirouchakis(2003) Research issues on product lifecycle management and information tracking using smart embedded systems. Adv Eng Inform 17:189-202. https://doi.org/10.1016/j.aei.2004.09.005

[32] M. Colledani, T. Tolio, A. Fischer, B. Iung, G. Lanza, R. Schmitt, J. Vancza(2014) Design and management of manufacturing systems for production quality. Cirp Ann-Manuf Techn 63:773-796. https://doi.org/10.1016/j.cirp.2014.05.002

[33] D. Kiritsis(2011) Closed-loop PLM for intelligent products in the era of the Internet of things. Comput Aided

Design 43:479-501. https://doi.org/10.1016/j.cad.2010.03.002
[34] H.B. Jun, D. Kiritsis, P. Xirouchakis(2007) Research issues on closed-loop PLM. Comput Ind 58:855-868. https://doi.org/10.1016/j.compind.2007.04.001
[35] J. Rios, J.C. Hernandez, M. Oliva, F. Mas(2015) Product Avatar as Digital Counterpart of a Physical Individual Product: Literature Review and Implications in an Aircraft. Adv Transdiscipl Eng 2:657-666. https://doi.org/10.3233/978-1-61499-544-9-657
[36] A. Cerrone, J. Hochhalter, G. Heber, A. Ingraffea(2014) On the Effects of Modeling As-Manufactured Geometry: Toward Digital Twin. Int J Aerospace Eng 2014. https://doi.org/10.1155/2014/439278
[37] A. Ladj, Z.Q. Wang, O. Meski, F. Belkadi, M. Ritou, C. Da Cunha(2021) A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective. J Manuf Syst 58:168-179. https://doi.org/10.1016/j.jmsy.2020.07.018
[38] K.Y.H. Lim, P. Zheng, C.H. Chen(2020) A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives. J Intell Manuf 31:1313-1337. https://doi.org/10.1007/s10845-019-01512-w
[39] R. Wagner, B. Schleich, B. Haefner, A. Kuhnle, S. Wartzack, G. Lanza(2019) Challenges and Potentials of Digital Twins and Industry 4.0 in Product Design and Production for High Performance Products. Proc Cirp 84:88-93. https://doi.org/10.1016/j.procir.2019.04.219
[40] C.L. Wu, Y.C. Zhou, M.V.P. Pessoa, Q.J. Peng, R.H. Tan(2021) Conceptual digital twin modeling based on an integrated five-dimensional framework and TRIZ function model. J Manuf Syst 58:79-93. https://doi.org/10.1016/j.jmsy.2020.07.006
[41] G.D. Shao, S. Jain, C. Laroque, L.H. Lee, P. Lendermann, O. Rose(2019) Digital Twin for Smart Manufacturing: The Simulation Aspect. Wint Simul C Proc:2085-2098.
[42] X.Z. Wang, Y.C. Wang, F. Tao, A. Liu(2021) New Paradigm of Data-Driven Smart Customisation through Digital Twin. J Manuf Syst 58:270-280. https://doi.org/10.1016/j.jmsy.2020.07.023
[43] K.Y.H. Lim, P. Zheng, C.H. Chen, L.H. Huang(2020) A digital twin-enhanced system for engineering product family design and optimization. J Manuf Syst 57:82-93. https://doi.org/10.1016/j.jmsy.2020.08.011
[44] T.Y. Lin, Z.X. Jia, C. Yang, Y.Y. Xiao, S.L. Lan, G.Q. Shi, B. Zeng, H.Y. Li(2021) Evolutionary digital twin: A new approach for intelligent industrial product development. Adv Eng Inform 47. https://doi.org/10.1016/j.aei.2020.101209
[45] X.J. Niu, M.L. Wang, S.F. Qin(2022) Product design lifecycle information model (PDLIM). Int J Adv Manuf Tech 118:2311-2337. https://doi.org/10.1007/s00170-021-07945-z
[46] J.S. Gero, U. Kannengiesser(2004) The situated function-behaviour-structure framework. Design Stud 25:373-391. https://doi.org/10.1016/j.destud.2003.10.010
[47] L. Hou, R.J. Jiao(2020) Data-informed inverse design by product usage information: a review, framework and outlook. J Intell Manuf 31:529-552. https://doi.org/10.1007/s10845-019-01463-2
[48] W. Zhang, S.J. Wang, L. Hou, R.J. Jiao(2021) Operating data-driven inverse design optimization for product usage personalization with an application to wheel loaders. J Ind Inf Integr 23. https://doi.org/10.1016/j.jii.2021.100212
[49] M.H. Cheng, C. Dang, D.M. Frangopol, M. Beer, X.X. Yuan(2022) Transfer prior knowledge from surrogate modelling: A meta-learning approach. Comput Struct 260. https://doi.org/ARTN 10671910.1016/j.compstruc.2021.106719
[50] E. Mysakova, M. Leps(2016) Press Weighted Average Surrogate: Trial Tests in 2d. Engineer Mechan:410-413.
[51] N. Quadrianto, Z. Ghahramani(2015) A Very Simple Safe-Bayesian Random Forest. Ieee T Pattern Anal 37:1297-1303. https://doi.org/10.1109/tpami.2014.2362751
[52] D.L. Sun, H.J. Wen, D.Z. Wang, J.H. Xu(2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362. https://doi.org/10.1016/j.geomorph.2020.107201
[53] E. Angel, D. Shreiner(2017) Application Development with WebGL. Sa'17: Siggraph Asia 2017 Courses.

https://doi.org/10.1145/3134472.3134481
[54] J. Jiang, G.F. Ding, J. Zhang, Y.S. Zou, S.F. Qin(2018) A Systematic Optimization Design Method for Complex Mechatronic Products Design and Development. Math Probl Eng 2018. https://doi.org/Artn 3159637 10.1155/2018/3159637
[55] M. Behzadian, S.K. Otaghsara, M. Yazdani, J. Ignatius(2012) A state-of the-art survey of TOPSIS applications. Expert Syst Appl 39:13051-13069. https://doi.org/10.1016/j.eswa.2012.05.056
[56] M. Grieves, J. Vickers(2017) Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. in: F.-J. Kahlen, S. Flumerfelt, A. Alves (Eds.) Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches, Springer International Publishing, Cham, pp. 85-113.