Digital Chemical Engineering 1 (2021) 100001

Contents lists available at ScienceDirect

Digital Chemical Engineering

journal homepage: www.elsevier.com/locate/dche

Original Article

A three-level hierachical framework for additive manufacturing

Yi Ming Ren[a], Yangyao Ding[a], Yichi Zhang[a], Panagiotis D. Christofides[a,b,*]

[a] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA
[b] Department of Electrical and Computer Engineering, University of California - Los Angeles, CA 90095-1592, USA

ARTICLE INFO

Keywords:
Computational fluid dynamics
Additive manufacturing
Machine learning
Machine data analytics
Edge-cloud interface
Industry 4.0

ABSTRACT

Metal alloy additive manufacturing (AM) has gained wide industrial interest in the past decade due to its capability to efficiently deliver complicated mechanical parts with high quality. However, due to a lack of understanding of the fundamental correlation between the operating conditions and build quality, the exploration of the optimal operating policy of the AM process is costly and difficult. In this work, a data-driven process optimization framework has been proposed for the additive manufacturing process, integrating machine learning, finite-element method (FEM) modeling, and cloud-edge data storage/transfer optimization. A three-level hierarchy of local machines, factory clouds, and a research center is introduced with each level responsible for its dedicated tasks. In addition, to ensure the efficiency of data transfer and storage, an edge-cloud data transfer scheme is constructed, which serves as a guideline for the data flow in the AM framework. Moreover, an overview of the connections between the proposed framework and the Industry 4.0 framework is presented.

1. Introduction

Additive manufacturing (AM) has been developed as a promising technique that has received wide utilization both in the prototyping and the production of parts in various industries. Compared to casting, forging, and other traditional manufacturing techniques, AM processes have many advantages, such as fast prototyping and manufacturing, high automation level, and versatile capability over various build geometries (Nandwana et al., 2016). In addition, a large variety of materials, such as polymers and metal alloys, can be used as the manufacturing media in the AM process (Gibson et al., 2014). The identification of the best operating recipes for these materials, especially for metal alloys, has been a major task for researchers in the field of AM. One of the most widely adopted methods for metal alloy AM is the laser powder bed fusion (LPBF) method (Frazier, 2014). Compared to other deposition methods, the LPBF method, by taking advantage of laser melting, produces build parts with better quality, higher resolution, and wider compatibility with build geometries. Thus, LPBF is the preferable technique for the manufacturing of devices that are quality-demanding and geometrically complex.

However, despite the advantages provided by LPBF, the exploration of the best operating policy for the production of different parts remains a difficult task. The difficulty can be attributed to the lack of an efficient method to understand the relationship between operating conditions and the build quality, as well as the absence of a systematic framework that takes full advantage of the massive amount of manufacturing data. Therefore, this work attempts to amend this gap by proposing an inclusive framework that takes advantage of machine manufacturing data, first-principles modeling and simulation results, and machine learning techniques. The framework is structured with a three-level hierarchy: the machine level, the factory cloud level, and the research center level. On the machine level, an automated monitoring and cross-validation scheme merging different process monitoring tools is proposed to enhance the robustness of in-situ defect detection using convolutional neural networks (CNN). If defects or build errors on the machine are detected, the build and error information will be sent to the second level: factory cloud. The factory cloud is a factory-level computer cluster with moderate computing power, which receives data from all local machines, where an operating policy update is performed based on a recipe update scheme to eliminate the defect in the subsequent builds. The recipe update scheme can be formulated using the combined effort of engineering knowledge and data-mining techniques such as recurrent neural network (RNN) models. In the cases where the recipe update scheme fails to provide an effective recipe update, a finite element method (FEM) simulation of the key physics of the AM process will be performed on the factory cloud to understand the physics and shed more insights on recipe updates.

In addition to the machine and factory cloud, the framework also proposes the inclusion of a research center cloud, which interacts with multiple factories, which serves as a data storage hub and is responsible for performing the higher computationally demanding tasks, such as the data-intensive derivation of new operating recipes as well as the

* Corresponding author at: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA.
E-mail address: pdc@seas.ucla.edu (P.D. Christofides).

https://doi.org/10.1016/j.dche.2021.100001
Received 21 June 2021; Received in revised form 17 July 2021; Accepted 17 July 2021
Available online xxx

formulation of FEM models for new build geometries. Moreover, the large amount of manufacturing data facilitates the design of advanced controllers based on machine learning (ML) methods, which is a topic that has received extensive research attention recently. For example, Bhadriraju et al. (2019) investigated the use of machine learning to perform adaptive model identification of a continuous stirred tank reactor (CSTR). On the other hand, Chaffart and Ricardez-Sandoval (2018), and Kimaev and Ricardez-Sandoval (2019, 2020a, 2020b) adopted neural network-based control schemes for thin-film growth processes. Also, Ding et al. (2021) proposed a design of a hybrid controller for the plasma-enhanced atomic layer deposition (PEALD) process by integrating proportional-integral (PI) and run-to-run (R2R) control schemes based on a recurrent neural network model. With such ML-based controllers, defect formation can be further minimized during production. Finally, along with the proposed framework, to optimize the data-transfer flow efficiency, a data-transfer scheme is proposed accounting for the severity of defects, data size, and flexibility. The AM framework as a whole is also an attempt to integrate the AM process into the industry 4.0 framework. Industry 4.0 is a novel concept that encourages the incorporation of smart manufacturing and information technologies into the manufacturing industry. Due to the close relationship between additive manufacturing and industry 4.0, this work also provides the context of industry 4.0 and how the proposed framework fits in the broad industry 4.0 framework.

2. AM framework hierarchy

The following sections discuss the proposed additive manufacturing framework, which is constituted of three levels: the machine level, the factory level and the research center level. An overview of the roles and responsibilities of each level is demonstrated in Fig. 1.

2.1. Machine

The lowest level of the additive manufacturing framework hierarchy is the machine level, where build parts are produced and manufacturing data are constantly generated and collected. An AM machine, e.g. EOS M290, receives the build recipe from the factory level and makes the build with the received recipe. A typical additive manufacturing recipe includes, but does not limit to, laser scanning path, hatch spacing, laser power, and powder thickness. Depending on the build geometry and material, the building parameters can vary dramatically, causing the building process to range from a couple of hours to a day. Due to a lack of an efficient real-time model that predicts the build part details online, it is crucial to monitor the AM process with the appropriate sensors to prevent and understand undesirable defects. The adoption of multiple sensors to provide better manufacturing results is prevalent in the additive manufacturing process. For example, Xu et al. (2018) developed a multi-sensor framework for the wire arc additive manufacturing, where individual sensors contribute differently to the final decision making. In addition, Dickins et al. (2018) also designed a multi-sensor in-situ monitoring system for the AM process. For the LPBF process, two of the most commonly used sensors to monitor the in-situ heat transfer aspect of the process, based on their functionalities, are the optical tomography (OT) sensor and melt pool (MP) sensor (EOS, 2018). The EOS M290 machine uses the EOS Suite, a process monitoring software, to perform the online monitoring using the information from these sensors. Both sensors can provide a unique perspective on the AM process. Specifically, the OT sensor outputs a whole image of the powder bed that summarizes the maximal temperature that occurred on the platform throughout the build process. The OT sensor is an off-axis sensor, which does not move along the laser path and in-sensor post-processing is necessary to adjust for the optical distortion due to different view angles. Additionally, since the OT sensor captures the entire build plate, the overall resolution will be lower. On the other hand, the MP sensor captures the local image of the melt pool. Since the MP sensor is an on-axis camera and aims to observe the transient behavior of the melt pool, it generates images at a much higher frequency than the OT sensor. Since the MP sensor focuses on a much smaller area, it produces a higher resolution over the local melt pool region. In other types of machines, e.g. Renishaw AM250, the EOS monitor suite is unavailable, but analogous cameras can be installed to achieve a similar effect. Specifically, (Lough et al., 2020) have installed a short-wave infrared (SWIR) camera on a Renishaw M250 machine to monitor the in-situ melt pool behavior of the process, similarly to the MP sensor. Experiments are also be done with thermal feature processing with the long-wave infrared (LWIR) camera to achieve similar results to the OT sensor.

2.1.1. Machine sensor monitoring

Even though sensor results are interpretable to process engineers, to further automate the defect identification process, it is desirable to adopt techniques that make the sensor results machine-readable. Moreover, the raw data size from sensors can be extremely large, reaching up to 30 Gb per build. Therefore, it is also important to perform efficient data reduction on the machine level to ensure the efficiency of data transfer speed and data storage size. Since AM sensor results are more easily interpreted as images, the raw time-series sensor data is first processed and converted into a layer-wise image of the build plate. For example, the thermal temperature data from an LWIR camera is transformed into a thermal mapping of the build plate through a thermal feature extraction algorithm. To classify these processed sensor images, CNNs have been widely adopted as an image classification technique and their accuracy has been acclaimed by a variety of industries (Krizhevsky et al., 2012). CNN can extract and recognize important features within an image in a highly efficient manner. Therefore, we propose to utilize CNN to process and interpret the AM sensor results. In this framework, each sensor is proposed to pair with a dedicated and specifically-trained CNN to perform the defect identification task on the local machine. The extracted features allow automated error detection and recipe update, which also allows redundant information to be discarded and keeps only essential information. It is also noteworthy that, a deployed CNN is relatively fast at execution and does not require a large amount of computation power to run. Thus, even if the machine-level computers may not have the strongest hardware, these computers are adequately equipped to perform the image classification task. Ren et al. (2021) has proposed a CNN defect detection workflow and tested using experimental data. The proposed CNN workflow can classify images at high accuracy and timely manner.

2.1.2. Sensor cross validation and machine transmission details

Different sensors, such as the OT, MP, or powder-bed sensors, provide different perspectives on the manufacturing process. Therefore, it is important to weigh the sensor results based on the intrinsic strengths of each sensor. Ren et al. (2020) has shown that the OT sensor performs better at identifying the over-melting problem, while the MP sensor performs better at identifying the under-melting problem. Similarly, the powder-bed sensor can also be cross-validated against the OT or MP sensors to identify problems such as recoater jams. Therefore, to fully take advantage of the strengths of the sensors, a cross-validation scheme is proposed to select the most reasonable sensor results using statistical analysis.

A crucial aspect of the proposed AM framework at the machine level is the information that is transmitted from machine to factory. CNNs and other process monitoring workflows can be implemented to reduce the amount of data transmitted. This filtered process information should be stored in an efficient and ready for use format before being transmitted to the next hierarchy of the framework, the factory level. In the following section, a cross-validation scheme between different sensors and the data transfer format is proposed for the example of two types of errors, e_1 and e_2, monitored by two different sensors, S_1 and S_2.

First, the problematic areas within each build are identified and their locations and error types are collected using the dedicated sensor CNNs,

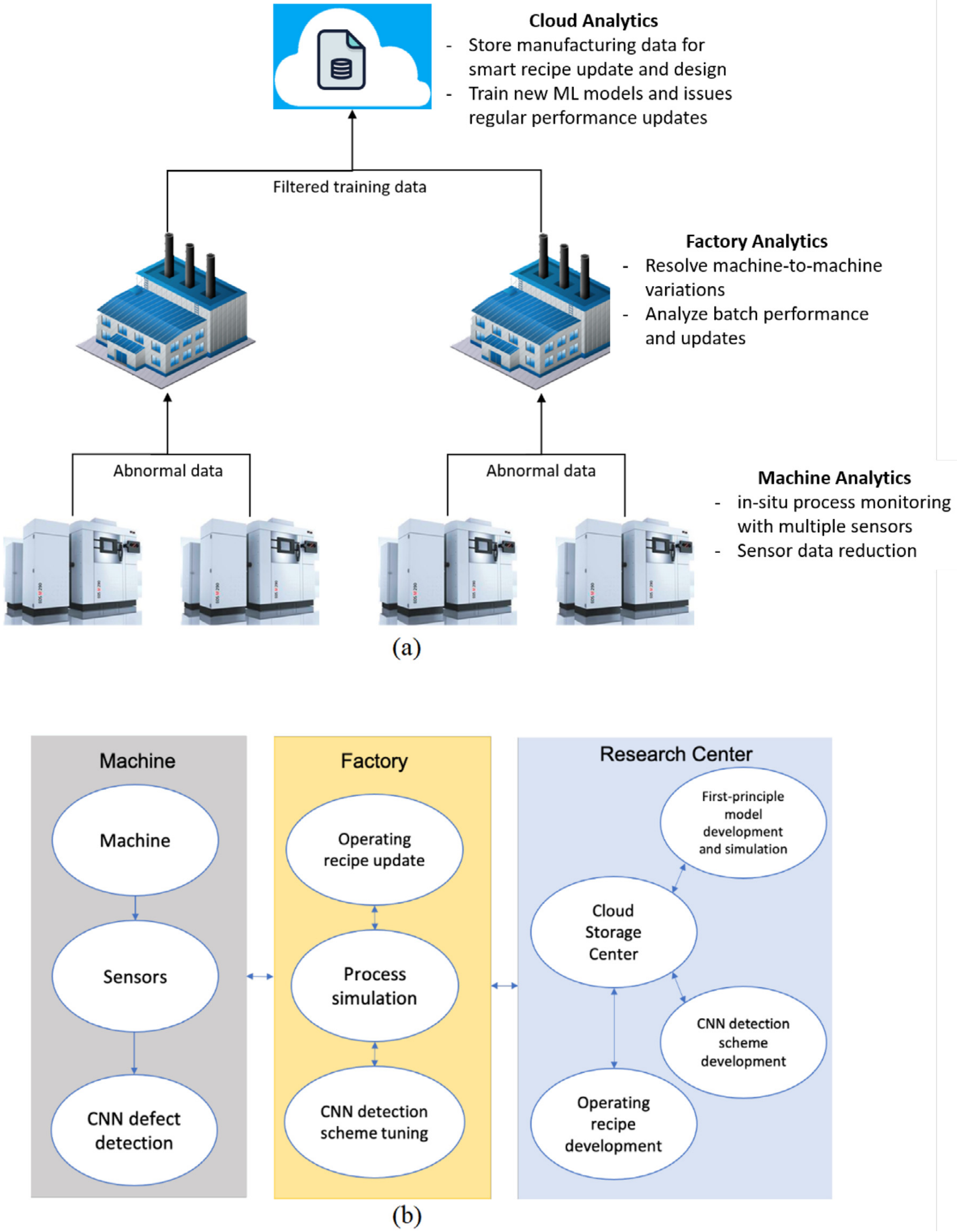

Fig. 1. An overview of the components of the additive manufacturing framework and their respective responsibilities. (a) A hierarchical depiction of the AM framework. (b) The machine level is shown in the gray block, the factory level is shown in the yellow block and the research center level is shown in the blue block. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$[(e_1, l)_j, (e_2, l)_j] = CNN_i(I_i), i \in \{S_1, S_2\} \tag{1}$$

where CNN_i is the trained network for a specific sensor i, I_i is the sensor raw data, i is the type of sensor image, e_j is the likelihood of a potential error label, l_j is the error location, $j \in [0, N_e]$ is the error index, and N_e is the total number of errors. Additionally, the output dimension is $N_e \times 2$ due to the two types of sensors involved in this case.

For one particular location l_j, it will have two types of error confidence levels reported by two types of sensors,

$$\begin{aligned}(e_1, l_j)_{final} &= \lambda_{S_1,1}(e_1, l_j)_{S_1} + \lambda_{S_2,1}(e_1, l_j)_{S_2}\\(e_2, l_j)_{final} &= \lambda_{S_1,2}(e_2, l_j)_{S_1} + \lambda_{S_2,2}(e_2, l_j)_{S_2}\end{aligned} \tag{2}$$

where $(e_i, l_j)_{final}$ is the overall likelihood of an error type i at location l_j and $\lambda_{k,i}$ is the weight determined by cross-validation for sensor k.

Then a filter is applied to every location to determine the most likely error. Confidence levels below a certain threshold will be regarded as error-free.

$$E = \left[filter\Big((e_1, l_j)_{final}, (e_2, l_j)_{final} \Big) \right] j \in [0, N_e'] \\ filter : \{\text{Pre-filter } e \text{ with a threshold; Then apply } \max(\cdot, \cdot)\} \tag{3}$$

where E is the cross-validated error list.

Although the example provided here uses two sensors and detects two error types, the proposed cross validation scheme can also be potentially generalized to more sensors and more types of errors. In addition, human knowledge can also be used for the determination of errors and these aspects can be explored in future works which focus more on the detailed implementation of the proposed framework.

2.2. Factory

As mentioned in the previous section, the factory level is a local computation cluster with moderate computing power and its physical location would ideally be in the AM factory to ensure efficient communication with the AM machines. The role of the factory level is to serve as a checkpoint and coordinator of the AM machines. When the pre-trained CNNs on the AM machine detect some potential defects, the error information, as well as the corresponding operating policies, will be sent to the factory for quality assurance and recipe correction if possible. Process engineers will be stationed at the factory level to perform quality control on incoming data. In the event of an error, process engineers first can utilize the local resources at the factory level including the local recipe book and simulation center in an attempt to find solutions to the error. Process engineers can transmit the error and its corresponding build information to the research center if further investigation is needed. In the case of misclassification of incoming data by the CNN algorithm, process engineers can selectively transmit the necessary data to the research center to improve and update the CNN algorithm.

2.2.1. CNN deployment monitoring

One of the most important tasks at the factory level is the quality assurance of the manufacturing process by skilled process engineers. At the machine level, it is not realistic to have process engineers monitor the various sensor results of each individual machine, hence the implementation of CNN algorithm to pre-filter out results. However, at the factory level, it is important for process engineers to regularly perform quality control on the machine products. While the CNN algorithm on each machine should already be performing at an acceptable high level, it still is susceptible to unexpected events and manufacturing data drift. For example, during the life cycle of an AM machine, the sensor calibration may shift and result in a systematic error in data distribution. If the new distribution is beyond the implemented range of the trained CNN network, then the CNN classification results will not be accurate and potentially impact the product quality. Examples of other unexpected events include sensor failures and faulty machines, which are also beyond the knowledge of the original CNN algorithm. Process engineers will monitor subsets of the incoming data from the machine level and respond to any potential events following the appropriate protocol. In addition, another benefit in having human experts monitor the results of the CNN data is the opportunity to improve CNN performance. If an incoming dataset contains incorrect classifications, the process engineers can selectively transmit a subset of the incoming images with the corrected labels to the research center for updates to the CNN algorithm. Detailed transmission schemes and a case study of this proposed workflow will be explained further in Section 3.

2.2.2. RNN workflow

On the factory level, a hybrid recipe update scheme is proposed which aims to provide smart recipe correction through the combined effort of both a high-level recurrent neural network (RNN) and human experts. A recurrent network is often used to predict outcomes based on historical data (Medsker and Jain, 2001). Since the machine-level manufacturing data are essentially time sequences of thermal and structural information, they can be processed by an RNN using a high-level analysis. Therefore, at the factory level, a pre-trained high-level RNN can be used as an operating recipe encyclopedia. The operating encyclopedia contains a large manufacturing dataset that can be used to derive the correlation between the operating conditions and potential defects. Therefore, the RNN can provide suggestions in the adjustment of AM operating policies when defects are detected. However, in the early phase of the manufacturing of a new build geometry, the manufacturing data are usually insufficient for the training of a meaningful RNN. Therefore, human knowledge can be extremely helpful in assisting the adjustment of AM recipes. In addition to that, human knowledge can also be applied to efficiently select the dataset to train the RNN with higher accuracy. However, with the collection of more manufacturing data, the feedback process can further be automated, thereby, reducing human intervention.

As mentioned in Section 2.1.2, the proposed recipe-update scheme is implemented using the transmitted error list, E, from individual local machines. Due to the standardized format of the transmitted data, the RNN-based workflow can be directly implemented without too much pre-processing. The formulation of the RNN network is shown as follows,

$$u' = RNN(E, b, u) \\ RNN : f(h(t), m(t), u(t), [e, l], b) = h(t+L), m(t+L), u'(t+L)$$

where h is a hidden state and m is a memory state. Using the error list and the additional build history around errors b, from an individual machine, the RNN can predict an updated recipe u' that can potentially fix the defects during the production. The same RNN network can be directly applied for multiple local machines and resolve the errors of the same type. It is noteworthy that, in the beginning phase of the manufacturing, it is difficult to derive a good RNN network due to the limited amount of manufacturing data. Therefore, human expert knowledge can also be adopted to facilitate the recipe enhancement.

2.2.3. Simulation center

The other role of the factory is to serve as a local AM process simulation center. The computation power of the factory cluster can be used to perform the solution of a pre-built FEM model simulation that does not require too much computational resource. These simulations, such as elementary structural and thermal analysis, can provide additional information regarding printing processes and avoid the usage of ex-situ analysis, which is often destructive to the printing product. Even with the limited computation power on the factory cloud, many types of simulations can be performed. Ren et al. (2020) has explored the possible simulations that can be performed on the factory level. According to the domain and scale that simulation models focus on, three types of simulations can be conducted on the factory level: micro-scale, meso-scale, and part-scale. An example of a micro-scale simulation is shown in Fig. 2(a). The shown 1-D microscopic transport model can be used to explore the thermodynamic and transport properties of the materials that are manufactured. In this particular case, the built 1-D model is used to compute the effective thermal conductivity of the powder material through numerical experiments. These types of simulations can be used as a useful supplement to the parameter determination studies, such as described in Economidou and Karalekas (2016). For meso-scale simulation, an example is shown in Fig. 2(b). The meso-scale simulation focuses on a portion of the built part, which can be a feature of interest or a sub-part. By constructing an FEM model, the local thermal history details of the part can be extracted for further analysis. The meso-scale simulation is particularly useful to obtain insights into the defects or distortions that occur on a specific geometric feature or location. Finally, part-scale simulation captures the dynamics of the entire

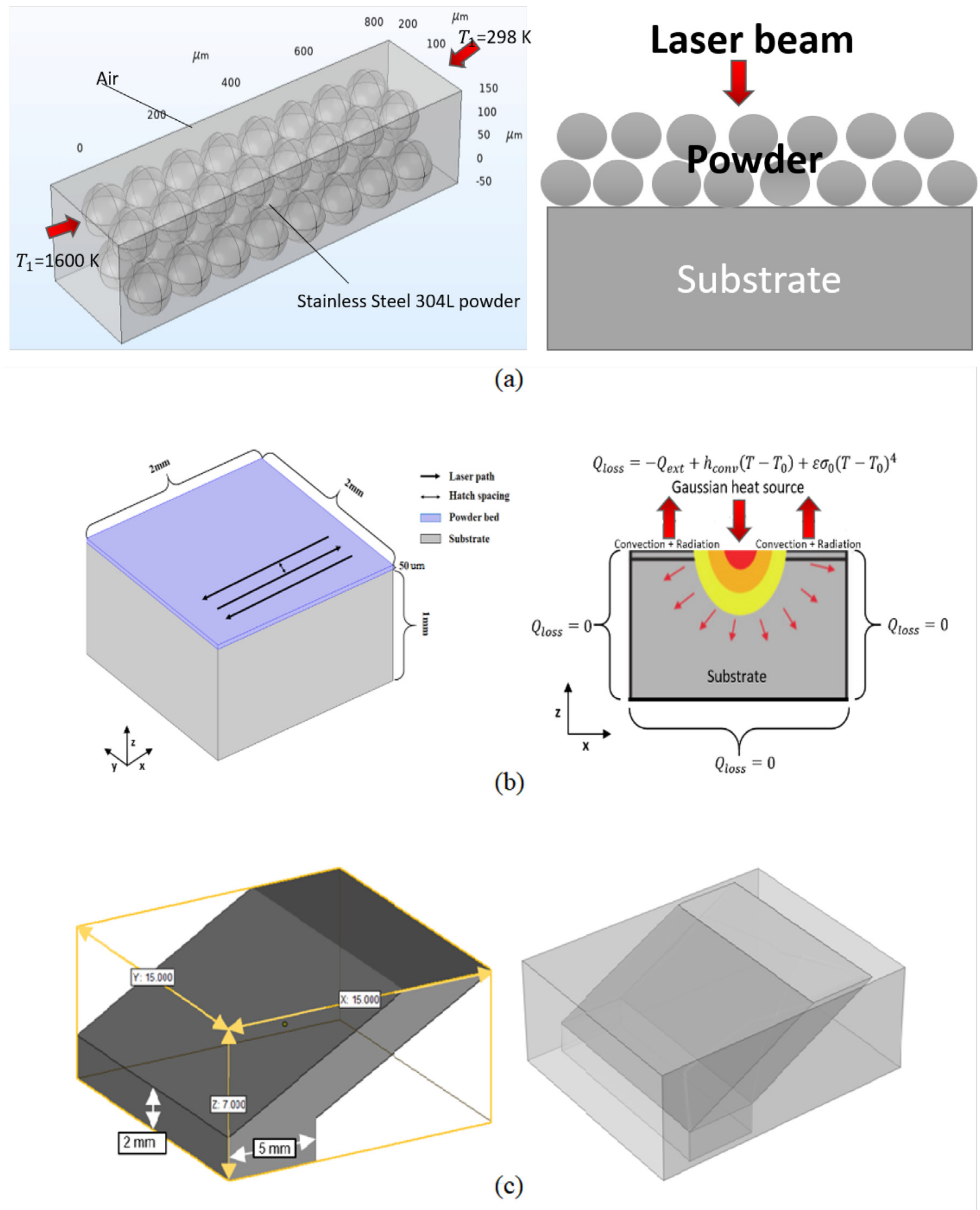

Fig. 2. Three examples of suitable simulations that can be run at the factory level. (a) Micro-scale simulation on the powder properties. (b) Meso-scale simulation on a portion of the build part. (c) Part-scale simulation of the entire build part.

build part or even a build plate where multiple parts are being built. An exemplary part-scale simulation geometry is shown in Fig. 2(c). The part-scale simulation can provide a holistic perspective of the building process, and it can be used to validate the building recipe and prevent part defects through simulation prediction. All three types of simulations are suitable for the computation resources available on a factory cluster. Additionally, due to the resemblance of part-scale simulation results with the sensor data, part-scale simulations can also be used as a data-augmentation method to further enhance the robustness of recipe design, which will be covered in more detail in the next section.

2.3. Research center

The highest level in the hierarchy of the framework is the research center, which can be assumed to be a supercomputer cluster equipped with strong computation power. Computation resources are typically limited on local venues due to the lack of sufficient hardware infrastruc-

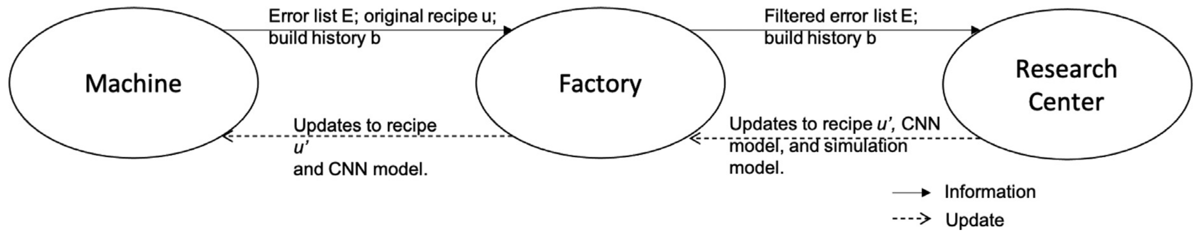

Fig. 3. An overview of the edge-cloud data transfer scheme of the additive manufacturing framework. The arrow orientations represent the directions of the data flow.

ture and high cost. Therefore, cloud computing has been proposed as an alternative to providing strong computational power to terminal users in a shared environment. The cloud computing market has witnessed significant growth in the past decade, and the technologies involved with cloud computing have matured to be able to meet the computation demand by the global research and commercial activities nowadays (Ren et al., 2015). Many cloud computing resources are available commercially, of which the most representative one is the AWS cloud service provided by Amazon (Amazon, 2015). In this framework, such a cloud computing hub is used as a dedicated research center for the AM process, which may serve and oversee hundreds of factories. With the strong computing power of the research center, tasks that are not feasible on the local machine and factory cloud can be performed here. It is noteworthy that the research center does not directly take part in the manufacturing process, and does not make recommendations for the recipe online. Instead, the research center serves as a knowledge hub that constantly takes in selected manufacturing data and performs machine learning studies, such as the training of a high-level recurrent neural network model, to provide the next-generation recipe for process optimization.

The research center can perform more computationally demanding simulations. For example, on the factory level, only one FEM model can be executed to validate a certain range of operating conditions due to the limitation of computational capacity. However, at the research center, a parametric sweep can be performed to inspect a large range of operating conditions, thus providing much more insights into the process. The type of simulations that can be performed at the research center also has a larger variety than at the factory cloud level. For instance, detailed simulation of the microscopic properties through the first-principles-based method can be performed, which is too computationally expensive on the factory cloud or local machine. For example, density functional theory calculations can be performed to obtain the necessary thermodynamic properties of the build material through the efficient approximate solution of the Schrödinger's equation (Gross and Dreizler, 2013). These first-principles-based simulation results, along with the large amount of manufacturing data provided by different machines on different build parts can provide sufficient quantity and generality of the additive manufacturing process. With these data, a robust model can be trained to provide insights on recipe designs for new build parts. Last but not least, another role of the research center is data storage. The stable infrastructure of supercomputer clusters makes the research center the ideal place for the storage and easy access of manufacturing data, which is typically on the scale of petabytes or even exabytes. However, as mentioned before, the research center is expected to receive data from hundreds and thousands of AM machines from the local factories. Therefore, the large amount of data influx and data transfer may be overwhelming even for a supercomputer like the research center. As a result, Section 3 proposes an edge-cloud data transfer scheme that can potentially optimize the data transfer and storage.

With the enormous amount of data gathered from different factories, the research center can also serve as a decision-making hub for the industry as a whole. Butt (2020) argued that enterprise-wide profitability is reliant on a number of integrated factors such as reliability, safety, flexibility, and environmental concerns. As discussed in Section 2.1.2, in-situ process monitoring of AM machines is an example of reliability and safety factors. The integration of ML techniques will allow more flexibility and turn-over rate on recipe design experiments through techniques such as AI-surrogate models and digital twins (Biegler et al., 2014). Environmental concerns may include the design of the input powder material and geometric shapes. It is important to allow consistent information flow between different factors as it contributes to aligning the enterprise-wide short-term milestones with long-term milestones. Following this approach, a number of short-term and long-term goals can be established with the large and expansive range of data collected in the research center.

3. Edge-cloud data transfer details

As mentioned in Section 2.3, due to the large size of the AM manufacturing data, it is crucial to perform data reduction and maximize the data transfer efficiency. The overall data transfer flow occurring in the AM workflow is shown in Fig. 3.

In summary, the data transfer procedure is proposed to be the following: first, an initial build recipe is provided to the AM machine by the factory cloud. As the machine builds the part, the sensor data, I, are processed with CNN, and an error list E is generated. The error list is further processed through the cross-validation scheme and a cross-validated error list, E, is generated. After the data processing has been completed, the data transfer is initiated from the local machine to both the factory cloud and the research center. The data size of the cross-validated error list, E, is estimated to be about several Megabytes per transfer and the data size of the raw sensor data, I, is estimated to be about several Gigabytes per transfer. Due to the large data size of the raw sensor data, not all data should be transmitted to the research center for every build. Instead, a data transfer frequency should be determined to ensure data completeness, data transfer bandwidth occupation, and data storage efficiency. To ensure the efficiency in data storage and transfer, the following data selection and transfer paradigm is proposed based on if a part defect or recipe error is observed:

```
Machine ->Factory
if error present then
    if new error then
        Factory ->Research Center at f1
    else
        Factory ->Research Center at f2 % if an old error is
        persistent
    end
else
    Factory ->Research Center at f3
end
where f1, f2, f3 are different data transfer frequencies;
f2 > f1 > f3
```

Specifically, in the proposed paradigm, the data transfers are all initiated in a linear fashion from the local machine to the factory cloud

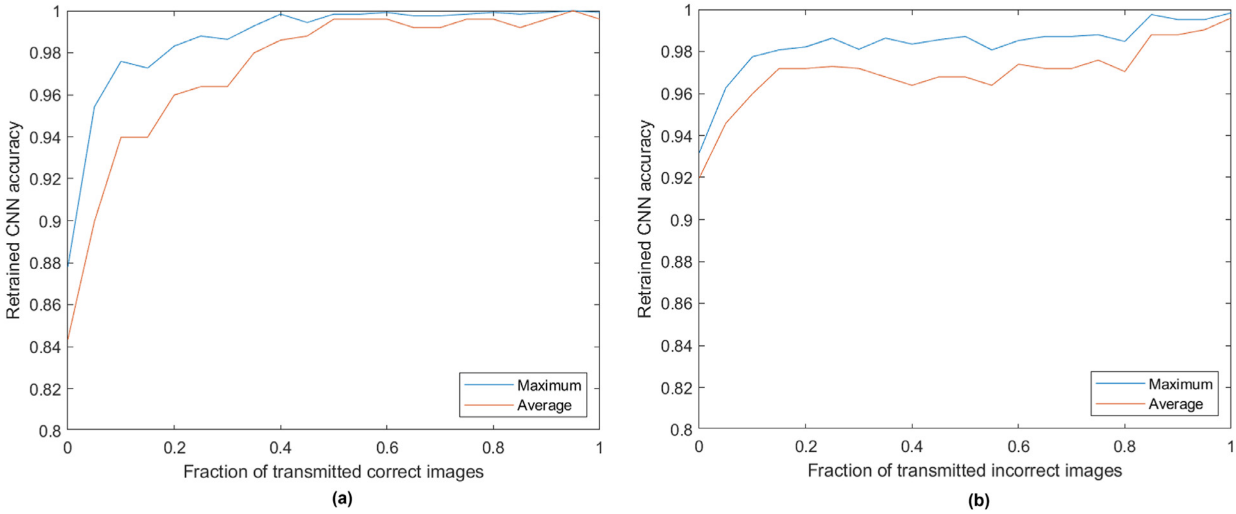

Fig. 4. Fraction of transmitted (a) correct and (b) incorrect images and their effect on the retrained CNN accuracy.

and finally to the research center to prevent data duplication. During manufacturing, if an error or defect is detected by the machine monitoring system and has not been encountered before, the machine will upload the manufacturing data to the factory cloud for error amelioration. After the data is sent to the factory cloud, according to a pre-trained recipe update scheme, an updated recipe will be deployed to the local machine, attempting to mitigate the error. Besides the feedback to the local machine, if the error has never been encountered before, i.e., a new error, the data will be also sent to the research center for bookkeeping and further study at a frequency of f_1. On the other hand, if the detected error or defect has occurred before and keeps reoccurring despite the recipe update from the factory cloud, the data will be sent to the research center from the factory cloud and the manufacturing should be halted if the severity of the defect is high. This data transfer will happen at a frequency of f_2. In the final case, suppose there is no error encountered during the manufacturing, the manufacturing data will still be transferred to the research center for routine backup. However, the frequency of routine backup, f_3, should be much slower than f_1 and f_2. The actual choice of data transfer frequencies is at plant designers' discretion, but an overall guideline for the priority of data transfer would be $f_2 > f_1 > f_3$.

3.1. Data transfer case study

Dataset used in this study is the same dataset used in the author's previous work (Ren et al., 2021). In summary, the dataset consists of processed LWIR camera data from different semi-arch geometries. The dataset is split into training, validation, and testing and the testing dataset is from a more recent build than the training and validation dataset. In a typical factory setting, the CNN classifier installed on each machine should already be performing at a relatively high level. The CNN used is able to achieve an accuracy of 92.3 % on a fairly balanced dataset of defective and non-defective images. The testing dataset will be used to represent a new set of build information from a typical AM machine. In this case study, the CNN classifier will be evaluated before and after transmitting the testing images and updating the CNN classifier.

In this case study, we will be investigating the effectiveness of the proposed framework, specifically, the transmission between factory and cloud. As mentioned at the factory level, the incoming images will first be classified in an AI-assisted manner in which process engineers will conduct quality control on the CNN classifications. First, we manually classify the new images to recreate the job of the process engineers who will monitor the incoming images. The new set of images are grouped into two categories: images of which the CNN-determined classifications that agree with human expert knowledge, and images of which the classifications that do not agree with human knowledge. As we are dealing with classifying unstructured data, images, the human classification, or the human-level performance, can be thought of as a good ground truth classification. Ideally, we want to minimize the total data transmission size while maximizing the effectiveness of the transmitted data. Therefore, these two categories of incoming images should be transmitted at different rates. Then, we will retrain the CNN classifier with both the new images and a subset of the original dataset. This is similar to the transmission from the factory to the cloud and the cloud retraining process. Finally, we will evaluate the performance of the newly updated CNN similarly to manufacturing ML deployment monitoring. The updated CNN is retrained five times to ensure performance consistency as there are slight variations between each training since the specific training images differ from training to training due to uniform selection. In addition, different weight initialization methods and stochastic gradient descent may result in slightly different results from time to time. Therefore, we reported both the maximum and average performance at each transmission rate in our result.

Before the retraining of the CNN, the CNN can achieve an accuracy of 91.3 %. This will be our baseline to which we will compare our updated CNN. As shown in Fig. 4 (a), we have varied the fraction of correctly classified images transmitted for retraining to investigate its effect on CNN performance. It is noteworthy that since we retrained our CNN with only 60 % of the original training dataset plus the new dataset, at fractions below 0.1, the updated CNN performs worse than the original. This decrease is to be expected since we did not use 40 % of the original training dataset causing some defect types to be missed. However, after merely using more than 0.1 of the new dataset, the updated CNN starts to perform much better. In our case study, the performance of the CNN stabilizes after around 0.4 of the total correctly classified images are transmitted for retraining. This supports our proposed framework of only needing to transmit a portion of all incoming images for retraining as the original CNN can already classify this subset of images correctly. However, it is still necessary to transmit them for retraining to keep up with the data drift and routine backup during manufacturing. In Fig. 4 (b), we have varied the fraction of incorrectly classified images to CNN performance. In our case study, the performance of the CNN continues to increase until around 0.85. This fraction is significantly larger than the previous 0.4 due to this subset being the incorrectly classified images. Specifically, we varied the fraction of correctly classified images first where all of the incorrect images are sent. After, we selected the best performing fraction, 0.4, and held it constant when varying the number of incorrectly classified images. We performed the study in this order because we know that correctly classified images contribute less than incorrectly classified images in the process of retraining. We did not vary both fractions at the same time since that increases the computation cost from $O(N + M)$ to $O(N * M)$. We do not need to transmit all of the images since some of the defect types are repeated, and sending 85 % of the incorrect images may be enough for the new CNN to recognize such a defect. In the case of a persistent error type throughout multiple iterations of the retraining process, this type of error should be transferred at a higher frequency or be assigned a higher weight when retraining the CNN. It is notable that the exact hyperparameters, such as transmission fractions and dataset ratios, are specific for our set of images and can vary when applied to another dataset. However, the general trends should be consistent throughout any dataset and in agreement with our proposed data transfer framework.

3.2. Job scheduling

The task of data transfer in AM highly resembles the work for the implementation of the aforementioned data transfer scheme. To address this task, one can take advantage of the various network load balancing schemes and job scheduling libraries available both open-source and commercially. Okwudire et al. (2018) investigated the usage of the user datagram protocol (UDP). For example, the transmission control protocol and the internet protocol ((TCP/IP) or open systems interconnection (OSI) in the field of networking is a great reference for the implementation of the data transfer scheme (Kozierok, 2005; Zimmermann, 1980). In the TCP/IP or OSI models, the data being transferred over the internet is divided into packets that are equal in size, which allows a convenient data integrity check and the usage of standardized processing protocol (Forouzan, 2002; Stevens and Wright, 1996). On the other hand, the sun grid engine (SGE) is an example of the job scheduling library. A job scheduler like SGE oversees a job scheduling queue and handles the job distribution and execution based on both job priority and resource availability. When performing the job scheduling, the job scheduler optimizes and balances the job load on all available computation nodes and ensures no node gets overcrowded or empty under various circumstances (Gentzsch, 2001). For the data transfer in the AM framework, due to the large data size of the manufacturing data, customization may be necessary when using existing software or libraries.

4. Additive manufacturing and industry 4.0

Additive manufacturing is often mentioned together with the concept of Industry 4.0, which represents the fourth industrial revolution. The key idea of Industry 4.0 is the integration of smart manufacturing and information technologies into the traditional manufacturing process, which is largely driven by the desire to combine digital and physical applications, to provide efficient product customization and to promote automation in manufacturing (Chen et al., 2018; Vaidya et al., 2018). Industry 4.0 typically features the usage and development of self-aware and self-learning machines, and cloud-based manufacturing and smart manufacturing are two of the main drivers of Industry 4.0. As a result, additive manufacturing is a key component of the Industry 4.0 due to its superior prototyping capability and close relationship with other components of the Industry 4.0 (Mehrpouya et al., 2019). Butt (2020) explained the interrelationships between additive manufacturing and cyber security, simulation, internet of things (IoT), and big data analytics.

Though still relatively not-well-developed, additive manufacturing is a broad field that has many sub-categories (Wong and Hernandez, 2012). Despite the large variety of materials that can be manufactured by AM, the materials that are most relevant with the framework and pathway of industrial 4.0 include metallic materials, smart materials (e.g., shape memory alloys and shape memory polymers), printable hydraulics and electronics and special materials & applications (e.g., jewelry, clothing and food) (Dilberoglu et al., 2017). Although the additive manufacturing data-driven framework proposed in this work focuses on the metal additive manufacturing and is based on the powder bed fusion (PBF) process, it can be tailored and adapted to the AM of other materials with the proper selection of sensors and potential transfer learning of the trained neural network. Moreover, the proposed data-transfer scheme is general-purpose and can be applied and customized for any additive manufacturing process that is data-intensive. In addition to the framework proposed in this work, there also exist other attempts that have been made to relate additive manufacturing with the industry 4.0 framework, from which this work draws inspiration. For example, Wang et al. (2019) has investigated the possibility of an IoT-enabled cloud-based additive manufacturing platform, where machine-learning techniques and hybrid human knowledge are used for the facilitation of rapid manufacturing. On the other hand, Buckner and Love (2012) has looked into the automation and the smart process-improvement of additive manufacturing through cloud computing and optimization. Moreover, recently Majeed et al. (2021) developed a big data-driven framework to assist the development of the smart AM process through big data utilization. Baumann and Roller (2017) also provided a comprehensive overview of the integration of cloud computing and additive manufacturing. Therefore, it can be seen that additive manufacturing is a key element in the industry 4.0 framework, which can potentially revolutionize the existing business models and manufacturing decisions (Garrett, 2014; Ivanova and Campbell, 2013; Lemu, 2018).

Finally, an emerging field within industry 4.0 is the integration and handling of cybersecurity concerns. When calculations, involved in the solution of process models, and/or when data analysis takes place in the factory cloud, the issue of cybersecurity should be carefully addressed as it is a problem that is occurring more and more in a variety of industries. To this end, cyber-attack detection and mitigation strategies should be implemented at the factory cloud to achieve early detection of potential cyber-attacks; a detailed treatment is outside the scope of the present work and the reader may refer to Wu and Christofides (2021)'s book for an in-depth study on the modeling, detection and mitigation of cyber-attacks.

5. Conclusion

In this work, a data-driven process optimization framework dedicated to the additive manufacturing (AM) process is proposed. The framework is based on a three-level hierarchy: machine level, factory level, and research center level. The machine level is the lowermost level, consisting of the AM machine or the computer equipped with the AM machine. The machine level is responsible for online data collection and real-time image processing. The second level is the factory level. The factory level is responsible for adjusting the recipe with a predefined recipe-update policy according to the defects reported by machines. In addition, the factory level can also coordinate data transfer and perform low computationally demanding simulation tasks to provide insights on the parts that are processed at the factory level. The research center level is used as the storage hub for the manufacturing data collected from the lower levels, i.e., factory and machine levels. Using the collected manufacturing data, the research center, consisting of a supercomputer with tremendous computational resources, can perform efficient and powerful data mining to improve and develop new process recipes. Moreover, the research center can perform first-principles-based simulations that further explore the material properties to facilitate the accurate characterization of the manufacturing process. In addition to the simulation/workflow framework proposed in this work, due to the large size typically involved with the AM manufacturing data, a data-transmission scheme is also proposed to serve as a data transfer guideline for the AM process, and a case study is presented with the data scale suitable for a typical AM application. Finally, the connection between AM and the

Industry 4.0 framework is elaborated to provide further motivation for the research in this field. Future work can be conducted to explore the efficient implementation of the framework proposed in this work and the integration between AM and Industry 4.0 concepts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Financial support from the Department of Energy is gratefully acknowledged.

References

Amazon, E., 2015. Amazon web services. Available in: http://aws.amazon.com/es/ec2/ (November 2012).

Baumann, F.W., Roller, D., 2017. Additive manufacturing, cloud-based 3D printing and associated services-overview. J. Manuf. Mater. Process. 1 (2), 15.

Bhadriraju, B., Narasingam, A., Kwon, J.S.-I., 2019. Machine learning-based adaptive model identification of systems: application to a chemical process. Chem. Eng. Res. Des. 152, 372–383.

Biegler, L.T., Lang, Y.-D., Lin, W., 2014. Multi-scale optimization for process systems engineering. Comput. Chem. Eng. 60, 17–30.

Buckner, M.A., Love, L.J., 2012. Automating and accelerating the additive manufacturing design process with multi-objective constrained evolutionary optimization and HPC/cloud computing. In: 2012 Future of Instrumentation International Workshop (FIIW) Proceedings. IEEE, Gatlinburg, TN, pp. 1–4.

Butt, J., 2020. Exploring the interrelationship between additive manufacturing and industry 4.0. Designs 4 (2), 13.

Chaffart, D., Ricardez-Sandoval, L.A., 2018. Optimization and control of a thin film growth process: a hybrid first principles/artificial neural network based multiscale modelling approach. Comput. Chem. Eng. 119, 465–479.

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., Yin, B., 2018. Smart factory of industry 4.0: key technologies, application case, and challenges. IEEE Access 6, 6505–6519.

Dickins, A., Widjanarko, T., Lawes, S., Stavroulakis, P., Leach, R., 2018. Design of a multi--sensor in-situ inspection system for additive manufacturing. ASPE and EUSPEN Summer Topical Meeting on Advancing Precision in Additive Manufacturing, Berkeley, CA. 248–152

Dilberoglu, U.M., Gharehpapagh, B., Yaman, U., Dolen, M., 2017. The role of additive manufacturing in the era of industry 4.0. Procedia Manuf. 11, 545–554.

Ding, Y., Zhang, Y., Chung, H.Y., Christofides, P.D., 2021. Machine learning-based modeling and operation of plasma-enhanced atomic layer deposition of hafnium oxide thin films. Comput. Chem. Eng. 144, 107148.

Economidou, S.N., Karalekas, D., 2016. Optical sensor-based measurements of thermal expansion coefficient in additive manufacturing. Polym. Test. 51, 117–121.

EOS, 2018. EOSTATE MeltPool: real-time process monitoring for EOS M 290. https://www.eos.info/software/monitoring-software/meltpool-monitoring(Accessed on 2020-07-19).

Forouzan, B.A., 2002. TCP/IP Protocol Suite. McGraw-Hill Higher Education.

Frazier, W.E., 2014. Metal additive manufacturing: a review. J. Mater. Eng. Perform. 23, 1917–1928.

Garrett, B., 2014. 3D printing: new economic paradigms and strategic shifts. Global Policy 5 (1), 70–75.

Gentzsch, W., 2001. Sun grid engine: towards creating a compute power grid. In: Proceeding First IEEE/ACM International Symposium on Cluster Computing and the Grid, Brisbane, Australia, pp. 35–36.

Gibson, I., Rosen, D., Stucker, B., Khorasani, M., 2014. Additive Manufacturing Technologies, vol. 17. Springer.

Gross, E.K., Dreizler, F.R.M., 2013. Density Functional Theory, vol. 337. Springer Science & Business Media.

Ivanova, O., Campbell, T., 2013. Additive manufacturing as a disruptive technology: Implications of three-dimensional printing. Technol. Innov. 15, 67–69.

Kimaev, G., Ricardez-Sandoval, L.A., 2019. Nonlinear model predictive control of a multiscale thin film deposition process using artificial neural networks. Chem. Eng. Sci. 207, 1230–1245.

Kimaev, G., Ricardez-Sandoval, L.A., 2020. Artificial neural network discrimination for parameter estimation and optimal product design of thin films manufactured by chemical vapor deposition. J. Phys. Chem. C 124, 18615–18627.

Kimaev, G., Ricardez-Sandoval, L.A., 2020. Artificial neural networks for dynamic optimization of stochastic multiscale systems subject to uncertainty. Chem. Eng. Res. Des. 161, 11–25.

Kozierok, C., 2005. The TCP/IP Guide: A Comprehensive, Illustrated Internet Protocols Reference. No Starch Press.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

Lemu, H.G., 2018. On opportunities and limitations of additive manufacturing technology for industry 4.0 era. In: International Workshop of Advanced Manufacturing and Automation. Springer, Changzhou, China, pp. 106–113.

Lough, C.S., Wang, X., Smith, C.C., Landers, R.G., Bristow, D.A., Drallmeier, J.A., Brown, B., Kinzel, E.C., 2020. Correlation of SWIR imaging with LPBF 304L stainless steel part properties. Addit. Manuf. 35, 101359.

Majeed, A., Zhang, Y., Ren, S., Lv, J., Peng, T., Waqar, S., Yin, E., 2021. A big data-driven framework for sustainable and smart additive manufacturing. Rob. Comput.-Integr. Manuf. 67, 102026.

Medsker, L.R., Jain, L., 2001. Recurrent Neural Networks: Design and Applications. CRC press.

Mehrpouya, M., Dehghanghadikolaei, A., Fotovvati, B., Vosooghnia, A., Emamian, S.S., Gisario, A., 2019. The potential of additive manufacturing in the smart factory industrial 4.0: a review. Appl. Sci. 9 (18), 3865.

Nandwana, P., Peter, W.H., Dehoff, R.R., Lowe, L.E., Kirka, M.M., Medina, F., Babu, S.S., 2016. Recyclability study on Inconel 718 and Ti-6Al-4V powders for use in electron beam melting. Metall. Mater. Trans. B 47, 754–762.

Okwudire, C.E., Huggi, S., Supe, S., Huang, C., Zeng, B., 2018. Low-level control of 3d printers from the cloud: a step toward 3D printer control as a service. Inventions 3 (3), 56.

Ren, L., Zhang, L., Tao, F., Zhao, C., Chai, X., Zhao, X., 2015. Cloud manufacturing: from concept to practice. Enterprise Inf. Syst. 9 (2), 186–209.

Ren, Y.M., Zhang, Y., Ding, Y., Liu, T., Lough, C.S., Leu, M.C., Kinzel, E.C., Christofides, P.D., 2021. Finite element modeling of direct metal laser solidification process: sensor data replication and use in defect detection and data reduction via machine learning. Chem. Eng. Res. Des. 171, 254–267.

Ren, Y.M., Zhang, Y., Ding, Y., Wang, Y., Christofides, P.D., 2020. Computational fluid dynamics-based in-situ sensor analytics of direct metal laser solidification process using machine learning. Comput. Chem. Eng. 143, 107069.

Stevens, W.R., Wright, G.R., 1996. TCP/IP Illustrated, vol. 2. Addison-Wesley.

Vaidya, S., Ambad, P., Bhosle, S., 2018. Industry 4.0 - a glimpse. Procedia Manuf. 20, 233–238.

Wang, Y., Lin, Y., Zhong, R.Y., Xu, X., 2019. IoT-enabled cloud-based additive manufacturing platform to support rapid product development. Int. J. Prod. Res. 57 (12), 3975–3991.

Wong, K.V., Hernandez, A., 2012. A review of additive manufacturing. Int. Scholarly Res. Notices 2012, 208760.

Wu, Z., Christofides, P.D., 2021. Process Operational Safety and Cybersecurity: A Feedback Control Approach. Springer.

Xu, F., Dhokia, V., Colegrove, P., McAndrew, A., Williams, S., Henstridge, A., Newman, S.T., 2018. Realisation of a multi-sensor framework for process monitoring of the wire arc additive manufacturing in producing Ti-6Al-4V parts. Int. J. Comput. Integr. Manuf. 31 (8), 785–798.

Zimmermann, H., 1980. OSI reference model - the ISO model of architecture for open systems interconnection. IEEE Trans. Commun. 28 (4), 425–432.