Automaatiopäivät24

Proceedings - Seminaarijulkaisut

ISBN: 13 978-952-5183-59-7

13-14 April 2021
Virtual Event

Edited by

David Hästbacka and Esko Juuso

ISSN: 1455-6502

SUOMEN AUTOMAATIOSEURA RY
FINNISH SOCIETY OF AUTOMATION

Tapani Raunio*, Ilpo Suominen, Santeri Myllynen, and Rasmus Karell

Machine Learning-based Requirements Classifier in the Analysis of Nuclear Power Specific Requirements

Abstract: Typical nuclear power plant projects include a huge volume of stakeholder requirements to manage. These include requirements that are hard to interpret and error-prone to analyze and allocate to correct technical disciplines and processes. By utilizing machine learning in the analysis of nuclear power plant requirements, designers' decision making in classification and allocation of requirements could be facilitated and thus, errors reduced.

Fortum has developed a machine learning-based requirements classifier utilizing recent advantages in natural language processing (NLP) and integrated it with a requirements management system. The classifier categorizes project specific requirements into pre-defined categories.

Utilizing pre-trained language models allows training of a classifier with 12 categories with less than 2000 labelled requirements. Our model achieves 98% accuracy in classifying requirements from similar document sources as used in the model training phase.

The success of the classifier encourages to investigate other potential areas to utilize NLP. These are, among others, atomizing (i.e., splitting up) long, especially multi-category requirements, classifying requirements written in Finnish, requirement fulfillment assessment, identifying similar requirements, etc.

Keywords: machine learning, NLP, requirements engineering

***Corresponding Author: Tapani Raunio:** Fortum, E-mail: tapani.e.raunio@fortum.com

Ilpo Suominen: Fortum, E-mail: ilpo.suominen@fortum.com

Santeri Myllynen: Fortum, E-mail: santeri.myllynen@fortum.com

Rasmus Karell: Fortum, E-mail: rasmus.karell@fortum.com

1 Introduction

All safety-critical projects need requirements management to demonstrate that the outputs meet the requirements that have been set. Requirements management and engineering is a challenge in nuclear industry projects due to the large amount of requirements. Solely the Finnish regulatory guidance includes several thousand requirements, and requires to justify and apply relevant standards in the design. Considering all the levels in the requirements hierarchy, the amount of requirements can be from thousands to tens of thousands, and their content vary from very specific to very generic.

Successful projects must track, elaborate, and manage requirements from several sources. Requirements need to be analyzed and allocated to correct technical disciplines and processes as they affect design choices and process. A single misallocated requirement can lead to design changes and affect multiple technical disciplines and processes. Changes to the design in late phase of a project usually carry a hefty price tag, and pose a substantial risk of delay.

Fortum has developed a requirements classifier based on utilizing recent deep learning model architectures and natural language processing (NLP) to support requirements allocation. The requirements classifier is capable of suggesting one or multiple technical disciplines and processes for the experts' consideration. Motivation is to ensure that relevant technical disciplines are identified in the requirement analysis phase.

In this paper we describe our recent developments and suggest further research topics. This work is built upon Fortum's earlier studies on the topic [1, 2].

2 Use of Requirement Categories

2.1 Use Case for Requirements Classifier

Requirements driven design together with requirements management ensures that the design should meet all stakeholder needs and expectations.

Requirements come from many different sources, which are written in different styles, contain overlapping requirements, and in some cases contain even contradictory requirements. A requirement can be hard to interpret, and when analyzing large amount of requirements, the humans' limited ability to concentrate on a specific task causes errors.

The later possible mistakes in the design of a product are noticed, the greater the additional costs and delays are. Correcting errors in the late stage of a project causes redesign, which leads to larger costs.

Improving requirements analysis ability decreases the probability of design errors. Our use case is to use NLP based requirements classifier as a support tool in requirement allocation.

2.2 Integration with Requirements Management

Requirements management can be done with a wide range of tools. In our case we use Polarion® ALM™ (later Polarion) Requirements Management System application.

We chose to integrate our requirements classifier with Polarion in order to automate the use of the classifier. From the user point of view adding a new requirement will lead to Polarion to suggest one or more requirement categories for the requirement. The final decision of correct requirement categories is always made by the user.

In the background Polarion and the classifier are running on separate servers, and communicating with web requests. In addition to the classifier server there is also a server to collect user saved requirement categories from Polarion, and a training server. Collecting categories saved by the user allows us to gather predicted and actual requirement categories. Training server is used to train new classifier models, and to analyze performance.

2.3 Requirement Categories

In this work we used 12 requirement categories, which consist of the following technical disciplines:
- Control Center Engineering
- Electrical Engineering
- HVAC Engineering
- I&C Engineering
- Process Engineering

and following process categories:
- Configuration Management
- Decommissioning
- Licensing
- Qualification
- Quality Management
- Requirements Management
- Verification & Validation

The selected requirement categories are based on our view of the main design life cycle related processes. At this point of development the categories do not cover all technical disciplines. For example mechanical, civil, and layout engineering among others are missing.

2.4 Data Collection

The goal of the data collection phase was to prepare a large set of requirements from multiple different sources for requirements classifier training and evaluation.

The collected dataset contains requirements from various sources such as Finnish Regulatory Guides on nuclear safety (YVL), IAEA requirements, ISO and IEC standards. Altogether requirements were collected from 40 different documents. All the requirements were in English.

Each requirement was allocated to one or more requirement categories mentioned in Section 2.3. Requirement allocation was done by discipline experts.

The data collection phase resulted in 2,819 requirements with each requirement in one or more categories from 12 possible categories. The median amount of requirements per category was 178.

3 Requirements Classifier

3.1 Natural Language Processing Text Classifier

Our requirements classifier is based on recent advances in NLP, and the readily available pre-trained state-of-the-art models. A pre-trained language model is already trained to "understand" text to a high degree, and we only need to teach the model classification task.

Pre-training enables the classifier training to focus on the requirement categories and enables classification when there is only a modest amount of labeled training data available.

The common approach in language model based NLP text classification is to transform input text to model's internal representation that provides features to the actual text classifier. Depending on the language model and the use case, the language model is fine-tuned as text classifier is trained or language model can be frozen as text classifier is trained.

Using a deep learning based language model allows fast model adaptation to new requirement data sets, because feature engineering is not needed.

3.2 Models

In this study the following representative language models are studied: BERT, DistilBERT, RoBERTa, XLM and XLNet [3-7].

BERT model has become the baseline for NLP projects and other chosen models expand or modify BERT various ways. DistilBERT provides comparable performance with improved speed and reduced resource consumption. RoBERTa uses vastly larger training set, Dynamic Masking Pattern and modified loss objective compared to BERT.

XLM model is similar to BERT but trained simultaneously on multiple languages leading to enhanced multilingual representations. XLNet uses generalized autoregressive approach with bidirectional context.

4 Experiments

We divided our 2,819 labelled requirements into following groups:

- Training dataset (n=1939)
- Validation dataset (n=414)
- Test dataset (n=390)
- Independent test dataset (n=76)

The training dataset was used to train the classifier part of the model, and fine-tune the pretrained language model for our classification task. The validation dataset was used in selecting model training hyperparameters. Model performance was assessed against the test dataset and the independent test dataset. The test dataset consists of requirements which were not used in the training or the validation dataset, but are from the same set of source documents. The independent test dataset consists of requirements from documents, which were not used in training or validation.

We used two different metrics to assess model performance: accuracy and match ratio. We define accuracy as the number of correctly labelled requirement categories divided by amount of requirements and requirement categories. Correctly predicted absences of categories are also counted in accuracy. In equation form

$$accuracy = \frac{\mathbf{1}}{\boldsymbol{n_r} \cdot \boldsymbol{n_c}} \sum_{s=1}^{n_r} \sum_{l=1}^{n_c} \mathbf{1}(\widehat{\boldsymbol{y}}_{s,l} = \boldsymbol{y}_{s,l})$$

where $\boldsymbol{n_r}$ is the amount of requirements and $\boldsymbol{n_c}$ is the number of categories, $\hat{y}$ is the ground truth and y is the prediction.

Match ratio is defined as the ratio of completely correctly predicted requirements to all requirements. A requirement is completely correctly predicted when all its categories (and their absences) were correctly predicted. In equation form

$$Match\ ratio = \frac{\mathbf{1}}{\boldsymbol{n_r}} \sum_{s=1}^{n_r} \mathbf{1}(\widehat{\boldsymbol{y}}_s = \boldsymbol{y}_s)$$

We trained and evaluated model performance in 5 different model architectures: BERT, DistilBERT, RoBERTa, XLNet, and XLM. In all cases the models predicted probabilities to all requirements classes, and 0.5 probability was used as a threshold for label prediction.

5 Results

The results for the test dataset are shown in Table 5-1. The performance in the test dataset is excellent with accuracy around 97-98% and match ratio in BERT based models 82-86%. All BERT based models are close to each other while XLNet achieving similar results. Only XLM has noticeable lower match ratio than other models. BERT achieves the best performance in the test dataset.

Table 5-1. Model results for the test dataset.

	Accuracy	**Match ratio**
BERT	0.983	0.869
DistilBERT	0.981	0.849
RoBERTa	0.978	0.821
XLM	0.972	0.734
XLNet	0.980	0.815

Model performances in the independent test dataset in Table 5-2 achieve good accuracy, but match ratio is poor compared to the performance in the test dataset. XLNet model performs best in the independent test dataset. Since the size of the independent test dataset is small, a bigger portion of performance and performance differences between models can be due to chance.

Table 5-2. Model results for the independent test dataset.

	Accuracy	Match ratio
BERT	0.906	0.355
DistilBERT	0.902	0.289
RoBERTa	0.914	0.394
XLM	0.898	0.158
XLNet	0.934	0.434

The results for BERT, DistilBERT and XLNet have been published earlier in [2], but were included here for comparison purposes.

6 Discussion

There are several possible ways to improve model performance. Collecting more high-quality data will likely bring performance improvements, as the models can be trained with a larger dataset. Different fine-tuning strategies could improve performance noticeably in the independent test dataset. One possible improvement strategy would be to add more context into the requirements themselves by adding for example the subheadings and headings structure to the requirement.

The main limitation of the current models is the need to retrain classifier when the amount of requirement categories or their content change. It is relatively easy to add new categories if they do not change already collected and labelled data. Introducing a new category, which might be present in the already collected data would require checking and relabeling all so far collected data.

There are many potential and exciting applications of NLP. In this work we utilized NLP to classify requirements. Possible future research topics include classifying requirements written in Finnish, requirements fulfillment assessment, atomizing complex requirements to multiple simple requirements, and identifying almost identical requirements. NLP could also be useful in analyzing the consistency and identifying possible contradictions in the document collection of the project.

7 Conclusions

We have shown that recent and easily available NLP models can be used in requirements classification task with excellent accuracy of 98% when requirements are written in a similar style as the requirements used in the training. Although our requirements classifier was trained with nuclear industry related requirements, similar results are likely reachable in other domains.

The availability of pre-trained NLP models and the ease of use of libraries have reduced the barrier for experimenting and developing new exciting applications. The amount of collected and classified data needed for model training and test data is reasonable.

Our requirements classifier is ready to be used as a support tool in requirements engineering. Integration with requirements management software allows to automate the use of the classifier, and to collect more data in real projects. More data will likely allow to train better classifiers in the future.

References

[1] Myllynen S., M.Sc. thesis, Aalto University, Espoo, 2019.

[2] Myllynen S., Suominen I. , Raunio T., Karell R., Lahtinen J., Developing and Implementing Ai-Based Classifier for Requirements Engineering, ASME J. of Nuclear Rad Sci., 2021

[3] Devlin J., Chang M., Lee K., Toutanova K., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL, 2019.

[4] Sanh V., Debut L., Chaumond J., Wolf T., DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, NeurIPS, 2019

[5] Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D. et al., RoBERTa: A Robustly Optimized BERT Pretraining Approach, arXiv, 2019

[6] Lample G., Conneau A. , Cross-lingual Language Model Pretraining, arXiv, 2019

[7] Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R., Le Q. V., XLNet: Generalized Autoregressive Pretraining for Language Understanding, NeurIPS, 2019

Terho Jussila*

Use of Symbolic Toolbox in Learning of Control

Abstract: Matlab, Simulink and Control System Toolbox have been used extensively in control engineering education. It is time to exploit symbolic computations more extensively than earlier for various reasons like more interesting pedagogic illustrations, increased motivation, verifications, saving time, better numerical stability and accuracy, reduced number of errors, student's self-training etc. In addition, some teachers feel that other Matlab tools hide some important challenges below their extreme flexibility and blackboxness. Therefore Symbolic Toolbox is here welcomed. It also seems to be suitable even for electronic open book exams, both alone and in co-use with other Matlab toolboxes. The paper discusses the strengths and the weaknesses of the symbolic computations.

Keywords: Teaching, Symbolic Mathematics, Symbolic Toolbox of Matlab.

***Corresponding Author: Terho Jussila:** Tampere University (TAU), E-mail: terho.jussila@tuni.fi

1 Introduction

Classical learning includes a lot of manual calculus and writing. It spends time to straightforward mechanical work with many dirty details faced in earlier studies. This reduces the time available for learning the new substance of the course. Manual work calls too easily for sloppiness errors like missing terms, wrong signs, erroneous variable identifiers etc. Correction of them may require a lot of extra time. A student may have faced the same troubles in a number of earlier courses and may become very frustrated to have the same pains again. This may damage her motivation.

Matlab and especially its Simulink extension, later SL, has been used for many illustrative simulations [1]. The presence of the SL block diagram e.g. helps the student to catch and keep in mind the wider context of the challenge. But SL can be used for only numerical simulations and only over a finite time range. In fact, all the parameters, input functions and initial conditions of the model should be available as floating point numbers. More-over, the effect of the choice of the integration algorithm and its accuracy settings to the accuracy of the results obtained is not always clear. In fact, many students are not really paying attention to topics like stiff systems, step size, variable step algorithms etc. Even a poor choice of Stop Time may lead to extremely poor conclusions: A simulation with too short a Stop Time may suggest stability of the system while a larger Stop Time would refer to instability. The claims presented can also be applied for other Matlab simulators like ode45 and dde23.

Control System Toolbox, later CST, can be used for simulation and other analysis of certain LTI (Linear Time-Invariant) systems. Even users of CST simulation tools should pay careful enough an attention to the selection of the finite Stop Time and the constant time increment to avoid misunderstandings. In addition, `lsim` is not able to simulate all the interesting forced responses correctly. Moreover, `lqrd` uses an unstable algorithm.

The author has felt since 1980s that optimization should be given more emphasis in the early education instead of teaching controller tunings based on a few simple arithmetic operations. Manual minimization/ maximization of may be tedious.

The weaknesses of manual, SL and CST work call for an additional toolbox to complement the old working practices. Below the use of Symbolic Toolbox, later ST, is recommended and defended. ST enables interesting fast enough illustrations and verifications, correct results more probably. ST may be used for solving analytically interesting problems too tough or boring for manual calculus. In many studies ST makes it possible to avoid early rounding and truncation errors. In some cases ST can provide numerical stability not easily available otherwise. The student may use ST in self-teaching by challenging her manual calculus with own-made problems for which the correct solutions are computed exploiting ST. Using ST in all phases including electronic exams may save totally a lot of time.

2 Ideas for a Basics on Analog Control

All the students are not familiar enough with basic commands, scripts and functions of Matlab when they arrive in the course. Most of them have zero experience on ST. This calls some very early Matlab training. The author offers part of that in an early voluntary PC work which gives the student a few extra points. The work may be supported by a video and pre-shared Matlab files. Matlab commands may be run from inside script. It is easy to find training examples which even bridge the student's earlier learnings and the new topics, strengthen the old skills.

Below possible use of ST is discussed by answering to questions like *what*, *why*, *how*? The hints, proposals etc. are organized according to the topics of a course on analog control. See e.g. [2] and [3] for references. Appendices give some examples uses of ST.

1. Introduction to Control: Basic use of ST is illustrated using the dependences of acceleration, velocity, maximum velocity and spatial displacement of 1D motion. In addition, the students are taught to program scripts and functions of Matlab. The demos support maximization and minimizations of some functions faced in later topics.

2. Mathematical Modelling: Equation solvers are applied for solving equilibrium values and for linear regressions of polynomial I/O models. This may include modelling of an actuator or conversion of values of e.g. displacements, pressures etc. onto 4…20 mA etc. using e.g. first order polynomials. In addition, simple enough differential equations may be solved using `dsolve` to challenge later solutions to be provided by e.g. Laplace Transforms or response theory of LTI state-space models.

3. Linearization: Nonlinear differential equations are replaced by approximate linear differential equations through linearization. This includes solving sets of static equations and deriving sample values of the appropriate partial derivative functions. The work may be difficult, frustrating and time consuming and produce easily errors for many realistic mathematical models like those of Balance Systems [2].

4. Laplace Transforms and Transfer Functions

TF or Transfer Function of a LTI system may be defined via three different approaches to be explained below. The approaches imply the same TF expression as long as one ignores the existence conditions.

OTF or **Operator TF** $G(p)$ is based on Derivative Operator p [4]. OTF is not nice enough in all ST work since the formal product of the OTF $G(p)$ and the time domain function f(t) is not commutative, i.e. a product like $G(p)f$(t) is not equal to f(t)$G(p)$.

ETF or **Eigenfunction Transfer Function** G(s) of a SI or Single Input system has a complex parameter s. ETF appears in the particular solution $G(s)exp(st)$ found possible to the test input $exp(st)$ [2]. Hence, given a single LTI I/O equation replace the input function with exp(st) and the output function with $Gexp(st)$, conclude G, and use it as $G(s)$. If the model refers to some intermediate response functions (like state variables) x_k other than the output then each x_k is guessed a particular solution of the form $G_k\, exp(st)$ after which G can be concluded by eliminating each G_k from the equations concluded. ETF is nice enough in ST work since $G(s)$ and $f(t)$ commute with each other, i.e. $G(s)f(t)$ is equal to $f(t)G(s)$ within the eigenfunction experiment.

LTF or Laplace Transfer Function based on LT or Laplace Transform LT. By a definition LTF of a causal SI system is the (Unilateral) LT of Unit Impulse Response UIR. ST offers the convenient functions `laplace` and `ilaplace` for LT and Inverse LT, respectively. LT is very flexible in modelling and computation of both the natural response of the LTI system and a number of interesting forced responses like impulse, step, ramp, eigenfunction and sine wave responses. Commutativity of multiplication for scalar LT objects is appreciated. The author prefers LT approach to the other TF approaches because of the observations made.

5. Simplification and Splitting of Block Diagrams

Complicated TF block diagrams can be simplified without techniques like *Mason Rule* by just posing a suitable set of linear equations for the unknown LTs involved and solving the unknowns. See Appendix for a demonstration. Here one may use OTFs, ETFs or LTFs.

Use of a polynomial model of a response function in stability analysis is less risky than use of a single TF for the same study. There are two reasons, one related to the ability to conclude a minimal model, another enabling extensive enough a pole study. First, quite many students fail to obtain the minimal (minimum realization) TF of a feedback system if the feedback path TF is not trivial enough. Consequently, some students would make wrong conclusions in later analysis of stability and steady state control error. Second, consider a system with the output y and the inputs r and d. Denote the LTs of y, r and d with $Y(s)$, $R(s)$ and $D(s)$, respectively. Assume a polynomial model like

$$s(s+1)Y(s) = sR(s) + (s+1)D(s) \qquad (1)$$

The roots of the polynomial multiplier $s(s+1)$ of

$Y(s)$ reveal lack of internal stability which here means lack of BIBO stability of the TF from d to y. But it is a common mistake to miss this diagnosis by forgetting to analyze that TF. The mistake can be avoided through analyzing the polynomial model. It may be constructed in the ETF or the LTF framework from the polynomial models of the sub-systems using a suitable form of *Cramer Rule* [5]: First, one should conclude from the polynomial models of the sub-systems into a square set of linear equations for the unknown LTs and present the set in the matrix form

$$Mv = w \qquad (2)$$

with a square matrix M and column vectors v, w. According to Cramer Rule the $k'th$ element v_k of v satisfies the equation

$$\det(M)v_k = \det(M_k) \qquad (3)$$

where M_k can be formed from M_k by overwriting its $k'th$ column with the column vector w. Here $\det(M)$ is the characteristic polynomial informative enough for the overall stability study of the output modelled using the LT v_k.

The block diagram rules may be used to the reverted direction to present a TF in terms of e.g. parallel connections and/or series connections of simpler TFs. Polynomial division may be used to present a biproper TF $G(s)$ as a sum of a Gain TF K and a strictly proper TF $Q(s)$, and the sum represents the parallel connection of K and $Q(s)$. The Gain K is the limit of the original TF for infinite s while Q is obtained by simplifying the difference $G(s) - K$. A reasonable series connection of a biproper TF may be enabled by a suitable factored form of the TF due to the poles of the TF and (in some case) the zeros of the TF. A PFE or Partial Fraction Expansion of the TF (`partfrac`) due to its poles may propose a parallel connection in the paths of which one may need series connections of several dynamic subsystems. Manual calculus needed may be painful and call for errors, especially in case non-real and/or multiple poles while ST work is quite trivial.

6. Introduction to Stability

Stability analysis is partly based on the poles of the TFs and/or the LTs of natural responses. There are two competing definitions of the poles and the zeros. A definition uses the roots of the appropriate poly-nomials or the non-polynomial functions of the dividend and the divider of the TF or LT [1]. Another definition is based on the limits the absolute value of the TF or LT has when the Laplace variable approaches the candidate pole or zero [5]. The definitions may lead to different sets of the poles if the tf is not minimal (minimum realization). The difference may cause confusion in deciding between stability or lack of stability if the modelling and the analysis enabled is too superficial. A nice example to illustrate the challenge is to study the TF of Moving Average Filter. It has even infinitely many poles, and the work needed to solve them manually is quite tough. A fellow teacher even regarded that task as too tough for the students.

Stability theorems can be illustrated with a number of Inverse LTs: when time is not burned for mechanical calculus both the student and the teacher have more time and energy for various enlightening illustrations.

7. Final Accuracy, Feedforward Control, Permanence

Final Accuracy. Final Values of various error functions like Actuating Error, Control Error, Measurement Error, Modelling Error (of Model Reduction) may be studied using FVT or Final Value Theorem of LT. Use of FVT is preceeded by some tedious modelling along the guidelines of the previous topics. So availability of ST commands is appreciated. Remark: The use of the usual error coefficients assume unit TF for the sensor.

FF or Feedforward Control: Here ST may be used to simplify the block diagram to enable derivation of the TF of the FF compensator. Then ST may be used to check the properness, causality and stability of the compensator and analyze the success of the compensation. Now the analysis is easy if this is trained and programmed well enough in Topic 6. Finally, ST could be used to simplify the compensator.

Permanence: The permanence of the steady-state response is analyzed in terms of differential sensitivity. Availability of ST for the boring partial differentiations challenges both the student and the teacher to make a more extensive analysis with a number of sensitivities.

8. Dynamics

Speed and damping of transients of converging step responses may be studied using computed expressions and plots of e.g. Peak Times, Overshoots, Undershoots, Integral of Error, Absolute Error, Squared Error etc. A performance index may be minimized in standard ways using the appropriate ST tools. Some students have appreciated such minimizations but disliked the routine mechanical work of manual calculus. One may extend such studies even to simple enough delay-in-loop systems [6].

Again, the efficiency of ST demonstrations allows additional interesting studies on the effect of the poles and the zeros of the TF to the time domain behaviour.

9. Introduction to LTI State-Space Models

The coefficient matrices of the state-space model can be formed as Jacobian matrices of the appropriate mappings (`jacobian`). The eigenvalues, the DC Gain and the TF of the model can be easily be found using ST.

Controllability and observability are studied using the easiest test matrices based on Cayley-Hamilton Theorem [5]. Even systems with more than two state variables may be easily studied with Matlab. Here the full accuracy of ST computations for matrices created with `sym` allows the correct diagnosis not always available in floating point computations.

The use of ST saves some time for Big Picture and a deeper discussion of the generic LTI theory including careful enough analysis of both external stability and internal stability.

10. Oscillations and Frequency Response (FR)

The need of complex numbers in typical frequency response studies make the topics quite difficult and aversive for many students. The pain can be helped by allowing ST work for the challenge. In addition, the separation of a sine or cosine response of a BIBO stable system into the sum of the decaying transient and the oscillating component can be easily demonstrated using either a convolution formula or Inverse LT. Moreover, typical plots of Nyquist, Bode and Nichols can be easily constructed. *A hint*: Replacing `abs(c)` with `sqrt(c.*conj(c))` may imply simpler expressions.

Command `help lsim` of CST claims that `lsim` may compute the response for an input like the sine function. But this is not true which the author sometimes illustrates for the students using ST. `lsim` fails even for e.g. a parabolic input function.

11. Frequency Response Studies of Stability

Ultimate Cycle Tuning Method of Ziegler and Nichols or ZN is interesting from pedagogic reasons while it may be regarded as naïve from engineering point of view. The method may be studied from various points of view for a few example feedback systems and/or just generically. The loop TF may include a gain, an integrator and a delay or be a 3rd order rational TF like that in R/W Head of a Hard Disk Drive problem [DB], see Appendix. One may compute Critical Gain, Critical Period, Gain Margin, Phase Margin and Delay Margin with ST and perhaps “verify” success via simulations. One may design Phase Margin and/or Gain Margin with ST commands. A generic study could find a bound of Stability Margin obtained for PID control due to ZN tuning. This is boring work manually, less boring with ST but demonstrates to the student some critical enough a thinking.

Course books devote some attention to the 2nd order critically stable zeroless Type 1 Loop TF to understand the speed and the damping of the closed loop system, in terms of the phase margin. One could also study the. Stability Margin implied. Its manual calculus is very awkward while ST can be used to find the answer.

Phase Margin design for stabilization of Double Integrator with Phase Lead Controller may also be studied using ST. Maximum Phase Shift available from the controller may be computed.

12. Frequency Response Studies of Dynamics

Pass Angular Frequency, Bandwidth, Peak Gain of e.g. Damping Ratio Model can be calculated manually with some effort while a student already familiar with the earlier theory and ST can compute and plot them quite rapidly and correctly.

Sensitivity of Closed Loop FR with respect to FR of Process or Sensor may be studied using ST. Generic studies are possible: think e.g. the sensitivity at the Cross-Over Angular Frequencies and the bounds of sensitivity at an AF equal to BW.

Spectral factorization problems may be solved using the poles and the zeros of the suitable functions.

13. Root Locus

Candidate break-away points and break-onto points of Re Axis not on the open loop poles or zeros can of course be found using the 1st order derivative function of the pole-to-gain mapping as pointed in many books [3]. Classifying the candidate points into the two different categories is often possible using the 2nd order derivative: a negative number would mean break-away, a positive number would mean break-onto, and a zero number would call additional information from a suitable higher order derivative. Again, finding the derivative functions may be painful but perhaps trivial enough using ST. Note that such a study is possible for even some non-rational Loop TFs.

3 On Digital Control

Many comments presented for analog control can be re-used here. Add a few comments for DT vs. digital control [7]:

First, manual calculus for digital control of a CT or Continuous-Time process is more cumbersome than that for analog control. A typical reason for this (beyond the simplest examples) is the larger number of non-zero model coefficients implied by the use of the hold circuit like ZOH. The difference implies system zeros even when the analog process has no zeros. Even this makes the calculus and theoretical reasoning more challenging. The stability tests (like Jury, Schur-Cohn) of digital control are more demanding than e.g. Routh Test of analog control. Use of ST may help in this pain.

A course on digital control may include more matrix studies. A TAU course actually devotes 1/3 of its width to Type 0 and Type 1 state feedback control, state observers and observer based control. Before it matrices are used e.g. for c2d conversions and TF computations. Manual matrix work may require too much time. So ST is appreciated again to save more time for the bigger picture.

In general, many Matlab tools available for double precision computations are numerically robust. However, `lqrd` of CST used for a hybrid study of optimal state feedback regulation is numerically ill-conditioned. In fact, it computes a matrix exponential for a matrix which has at least an eigenvalue right to Imaginary Axis if the LTI state process has at least an internal feedback. Actually, the tool is poor for any other system than the pure integrator system. Usually, this is not a major problem since one typically uses small enough a Hold & Sample Time h for which the problem is minor or ignorable. But for a stiff system somebody might select a "large" h for which some loss of accuracy could be possible. Fortunately, ST can help in the problem: For some low order processes `int` (`symbolic/int`) and `expm` can be applied for numerically stable com-putations. For numerical model matrices joint use of the eigendecomposition method of `expmdemo3` and `int` is possible if the eigenvector matrix is non-singular.

Putting the observations together, ST seems to be even more useful in digital control than in analog control.

4 ST in Open Boox Exams

Currently, TAU campuses enable electronic exams in camera-supervised EXAM classes. In the class the students have access to Matlab and many toolboxes including their electronic documentations. The author has organized/organizes there Open Book Sub-Exams. A sub-exam may cover 1-4 week topics. The students is asked to combine manual work and Matlab work. The student has access to the pdf and Powerpoint files (.pdf, .pptx) of the lectures and formulae pages but not to any exercise files. Skills for basic Matlab and Simulink are needed. Use of ST is allowed while the right to use of CST may be limited.

A problem may include both symbolic parameters and double precision parameters. The student may solve the problem manually, with ST or using both manual work and ST to handle just the most difficult and lengthy math reasonings. In some cases she is adviced to select the symbolic parameters example numerical values and to challenge her answer by the proposals of other Matlab ways like ordinary commands, Simulink or CST. To save her time for the main job a Simulink block diagram or a semi-ready Matlab script may be given for the purpose. The students is encouraged to maximize the use of ST to save time by avoiding the manual work which can take a lot of time: the interface of EXAM software does not allow very quick writing. Copy & Paste can save time, a simple command can do a major job, and one may avoid the usual errors with signs, missing terms, miss-written terms etc.

Appendix 3 shows an example problem in which the actual work is made with ST while the result may be "verified" with a Simulink simulation.

5 Discussion & Summary

ST deserves some attention due to various reasons. First, spending less time in usual mechanical mathematical reasoning saves more time for catching the new substance of the course. One may save time in contact teaching, homework and exams. Currently, COVID rules have limited the exam time of EXAM classes to 115 minutes which means a modest time pressure in a sub-exam of 4 week topics. After good enough a training a user of ST can save 5-10 minutes in a topic. This is partly enabled by the use of Copy & Paste followed by appropriate editing and the possibility to return the answers as Matlab script files (including Live Scripts). Second, ST can also avoid some sloppiness errors typically made by the students. Third, ST work can eliminate rounding errors and truncation errors in many cases. The worst possible failures due to numerical instabilities may perhaps also be avoided by appropriate use of ST. Fourth, the student may use ST in her training phase to provide the correct solutions of the benchmark problems created by herself. Fifth, as seen already earlier when use of Simulink and CST began, a new CAD tool may increase the student's interest on the course topics. A student may have been frustrated with manual calculus in earlier courses and may therefore welcome an alternative.

There are problems with multiple if not even infinitely many solutions. It is a quite typical to find just one of the solutions in manual calculus. The same can occur in ST work if the best possible commands are not used. An appendix demonstrates appropriate use of `solve` (of ST) to find all the infinitely many solutions of a characteristic equation.

Naturally, there may be problems in the use of ST. This is especially true in the first course which uses ST if the student has no a'priori skills in ST beyond the courses. However, the skills acquired in it will make application of ST in later courses more fluent. In addition, too powerful ST commanding may also hide information.

The student should in some cases be adviced to answer in an algorithmic way instead of expressing the answers as explicit functions of the parameters referred in the question. It is not relevant to print an explicit solution of a problem into 2-3 A4 into an appendix of e.g. a report (as seen in some theses). Avoiding this by an algorithmic answer requires good design of the ST commands to which the student should be trained.

The student should also be asked to include comment lines in the scripts. Actually, she could write them in the very beginning to guide her commanding. Even in another respects one should follow a few good practices of software production.

Appendices

Three ST demos are included below to illustrate the powerfulness of symbolic computations:

Demo1. TFs of a Feedback System

```
% TFs of a Negative Feedback System.
% LTs of Process Output y,
% Set Point r, Measurement m,
% Actuating Error e, Disturbance d,
% Manipulated Variable u,:
syms Y R M E U D
% TFs of Controller,Process, Sensor:
syms F G H
% Solve Y, U, M, E:
so = solve(Y == G*U + D, U == F*E, ...
E == R - M, M == H*Y,Y,U,E,M)
Y = so.Y , U = so.U,
E = so.E , M = so.M
% Y = Gyr*R + Gyd*D , U = Gur*R + Gud*D
Gyr = diff(Y,R), Gyd = diff(Y,D)
Gur = diff(U,R), Gud = diff(U,D)
```

Demo 2. Moving Average Filter

ETF, zeros, poles and DC Gain of a Moving Average Filter with Data Window Length T. An interesting example the poles, zeros and stability have confused many students. TF can propose a block diagram.

```
syms T t positive
syms s
% Eigenfunction Input:
u = exp(s*t)
% Particular Solution:
y = int(u,t,t-T,t)/T
% ETF :
G = simplify(y/u)
pretty(G)
solve(G,s,'ReturnConditions',true)
so = ans
pa = so.parameters
co = so.conditions
TFzeros = so.s
solve(1/G,s,'ReturnConditions',true)
so = ans %
TFPoles = so.s
DCGain = limit(G,s,0)
```

Demo 3. Ziegler-Nichols Experiment

Critical Gain pcr and Critical Period Tcr of Ziegler and Nichols in Proportional Position Control of a DC Motor of a Hard-Disk Drive assuming an Ideal Sensor

```
syms s
syms L R K J b p positive
% TF of DC Motor [3]:
G = K/(L*s+R)/(J*s+b)/s
% TF of P Controller for P Gain p:
F = p
% Sensor TF:
H = 1
% Open Loop TF:
Q = G*F*H
% Angular Frequency
syms w real
assume(w >= 0)
% Frequency Response:
Qfr = subs(Q,s,i*w)
% Critical Angular Frequency:
wcr = solve(imag(Qfr) == 0,w)
% Critical Period
Tcr = 2*sym(pi)/wcr
% FR at wcr:
Qfr = subs(Qfr,w,wcr)
Qfr = simplify(Qfr)
% Critical Gain from
% Barkhausen Criterium:
pcr = solve(Qfr == -1,p)
pretty(pcr)
% Substitute Parameter Values [3]:
pcr = subs(pcr,{L,R,K,J,b},…
      {1e-3,1,5,1,20})
Tcr = subs(Tcr,{L,R,K,J,b},…
      {1e-3,1,5,1,20})
pretty(Tcr)
% From ST Data Type
% to Double Precision Numbers:
pcr = double(pcr)
Tcr_ = double(Tcr)
```

References

[1] https://www.mathworks.com/

[2] Åström K. J. , Murray R. M., Feedback Systems. An Introduction for Scientists and Engineers, 2nd edition, Princeton University Press, 2020.

[3] Dorf R. C., Bishop R. H., Modern Control Systems, 13th Edition, Pearson Education Limited, 2017.

[4] Rotella F., Zambettakis *I.,* An automatic control course without the Laplace transform An operational point of view, 2008, e-STA, 5 (3). 18-28. ISSN 1954-3522.

[5] James G., Advanced Modern Engineering Mathematics, Prentice-Hall 2003.

[6] Marshall J. E., Gorecki H., Korytowski A., Walton K., Time-Delay Systems: Stability and Performance Criteria With Applications, Prentice-Hall 1992.

[7] Franklin G. F., Powel J. D., Workman M., Digital Control of Dynamic Systems, 3rd Edition, Addison-Wesley 1998.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Timo Viitaniemi * and Henri Pettinen

Utilization of ROS2 with Edge Computing

Abstract: Utilization of distributed computing on industrial vehicle onboard systems, the edge devices, opens many possibilities to save development resources and time yet giving more flexible software design. Robot Operating System 2 (ROS2) is a prominent candidate for achieving these goals. It has raised great interest all around the world among researchers and big industrial players. In this paper we are exploring ROS2 from the edge controller point of view trying to answer to the following questions. How ROS2 changes the edge device application development? How ROS2 communication meets the real-time and security requirements? How portable and reusable ROS2 application components are? Evaluation is based on public research papers, articles, standards, and ROS2 documentation, but also on practical evaluation with ROS2 on edge devices. As part of this paper we implemented performance study on message size and sending frequency, data transfer latency, node's phase-wise startup latencies, and topic publisher's CPU consumption. We have identified that the ROS2 QoS and security settings do what they promise, but they should be used with care and with case by case consideration as careless use may bring undesirable slowness to the system or even loss of data.

Keywords: Robot Operating System 2, Data Distribution Service, Quality of Service, Edge Computing

***Corresponding Author: Timo Viitaniemi :** CrossControl Oy, E-mail: timo.viitaniemi@crosscontrol.com
Henri Pettinen: CrossControl Oy, E-mail: henri.pettinen@crosscontrol.com

1 Introduction

It is common on the machine control systems on the edge to combine different types of devices like displays, I/O controllers, Machine Learning (ML) computing accelerators, different types of sensors, cameras, etc. into single solution. That means there is need for a unified software platform helping the ecosystem to work seamlessly together.

One such approach to handle distributed computing on the edge control systems is the Robot Operating System (ROS). It got its start in 2006 in Stanford University's Personal Robotics Program [1] of two Stanford students Keenan Wyrobek and Eric Berger. The work continued by company called Willow Garage which developed ROS further around six years until the company was closed. Since then the ROS development has been done by Open Source Robotics Foundation which now days is called Open Robotics[2]. [3]

Although ROS has long history and it has been very popular with robotics, the first generation ROS1 had several shortcomings preventing it from industrial use:

- Single point of failure. Master node is crucial for all other ROS nodes to work.
- Lack of security as there is no encrypting in ROS1 nework communication.
- Lack of real-time support in communication protocol.

It is only during last few years, after the release of ROS2 in late 2017, it has been an alternative for industrial use.

It seems that ROS2 has raised great interest all around the world among researchers and developers. Here are some statistics ([4] , July 2020) about visitors to wiki.ros.org:

- China 19 %
- USA 16 %
- Japan 8%
- Germany 7 %
- South Korea 6 %
- India 5 %

Since 2019 there is increase of

- 87 % in ROS debian packages downloads
- 43 % in rosdistro github repository activities
- 26 % in number of research papers

This paper introduces the results from CrossControl's study project [5] upon utilization of ROS2 in industrial edge devices.

Organization: this paper is organized as follows

1. Introduction
2. Robot Operating System 2
3. Data Distribution Service
4. Quality of Service
5. Security
6. Evaluation
7. Conclusions

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

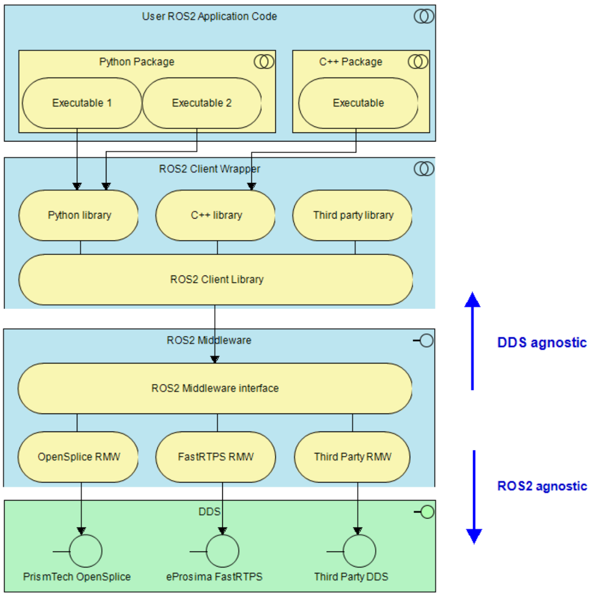


Fig. 1. ROS2 architecture.

2 Robot Operating System (ROS)

The name "Robot Operating System" misleadingly suggests that ROS is an Operating System (OS), but actually it is a set of software libraries and tools for building robotic applications. ROS attempts to provide large-scale software (SW) integration tools for robotic environment where is vast number of different SW and hardware (HW) components. [6]

The situation is similar to machine control systems on the edge, where controllers are communicating via standard interfaces like Ethernet and CAN bus. What ROS brings into this picture is that SW processes can communicate directly with each other over the Ethernet regardless of the physical device they are located in.

Fig. 1 illustrates ROS2 application architecture. User code accesses ROS2 client library to implement ROS2 features. ROS2 middleware is an abstraction of the communication channel.

The current official ROS2 release supports Python and C++ client libraries, but there are third party libraries also for C, Java, Node.js, and TypeScript . Typically Python is good for fast prototyping and C++ when performance is needed.

2.1 Packages and Executables

ROS2 application code is divided into packages which can be considered as containers for ROS2 code. Packages can be also used to share ROS2 application code with others and there are many free open-source packages available for ROS2 [7].

ROS2 executable is a code entry inside a package. Together with the package name, executable is used to start a specific ROS2 process.

When ROS2 package development is ready, it is built using universal build tool [8] that generates binaries and installation files to be used on the target device.

2.2 Nodes

When ROS2 package's executable is started on target device, a new process is created. This new process is called ROS2 node. Term "node" comes from the run time visualization graph where the processes are graph nodes and the peer-to-peer links arcs [6].

Nodes communicate with each other via ROS2 middleware interface and they may reside on different application processes or on different computers.

ROS2 nodes communicate with each other using specific communication patters called topics, services, actions and parameters.

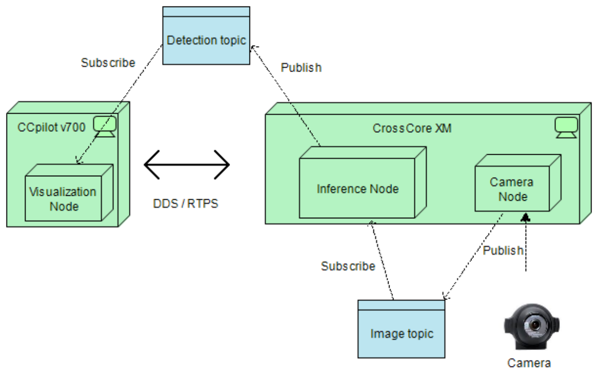


Fig. 2. Displaying camera image ML detection results on CCpilot v700 [9] display using ROS2 topics.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

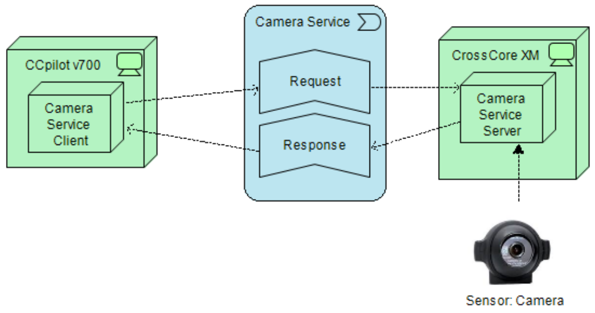


Fig. 3. Displaying camera image on CCpilot v700 [9] display using ROS2 service.

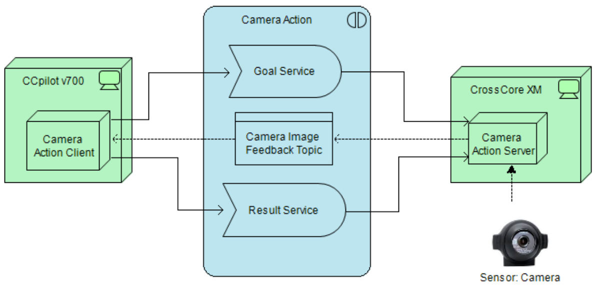


Fig. 4. Displaying camera image on CCpilot v700 [9] display using ROS2 action.

2.3 Topics

Topics work in publish-subscribe model so that publisher feeds topic data messages to DDS Global Address Space from where the subscribers receive the topic message data. In practice, ROS2 middleware takes care of all communication between publisher, subscriber and DDS so that code in publisher node only needs to call the publish method and subscriber node register a method for receiving the topic data. There can be one or more publishers and subscribers per topic on different nodes. The idea of topic has been illustrated with camera image and ML detection example in Fig. 2.

2.4 Services

The idea of services is illustrated with camera image example in Fig 3. Services work in call-and-response model where the service client requests a value from service and service server responses to the request with the requested value. ROS2 middleware takes care of all the communication between nodes.

2.5 Actions

Actions are used for long-running processes like when robot is navigating to a specific destination. They are constructed from two services and a topic. Action starts when action client initiates the action by action server's Goal Service. Then action server starts sending Action Feedback topic to the action client. Action client can stop the action server by using action server's Result Service. An example action is illustrated in Fig 4 where camera image feed starts when the client sets the Action Goal and ends when the client calls Result Service. The camera image is displayed in the v700 display while the action is running.

2.6 Parameters

Parameters are used to configure ROS2 nodes and are implemented internally as a service.

Each node is responsible for storing its own parameters. It is possible to get and set parameters outside the node itself allowing external behaviour-control of the node.

2.7 Interfaces

ROS2 interfaces are message-definitions for node communication. Each interface is defined in Interface Definition Language (IDL) file. An interface can be uniquely identified using ROS2 package and IDL file name. ROS2 IDL files are used for auto-generation of message code for several programming languages. This means developer only needs to define IDL file and start using it from the code.

Interface file contains all the fields for defining a message. Each field is defined by field type and name. This makes ROS2 messages type-safe, preventing errors, and be more user friendly. Field types correspond to the primitive programming language types [10]. It is possible to define default values for the fields and set fields as constants.

Complex (custom) field types can be defined by using other IDL files as a field type. This makes ROS interfaces flexible and promote type re-usability.

There are separate IDL files for topics (.msg), services (.srv), and actions (.action) since they all differ in communication. Topic communication has single part, services have two parts, and actions have three parts, according to their communication models. Each com-

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

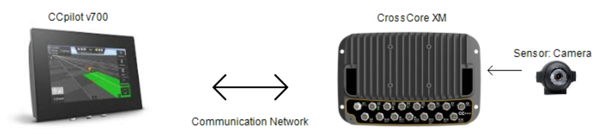


Fig. 5. Displaying camera image on CCpilot v700 [9] display in real life.

munication part has its own section in a message file that separate its field definitions from others. The separator between the parts are three dashes ("- - -").

When using interfaces at run time in node communication via ROS2 middleware, ROS2 IDL needs to be converted to DDS IDL [11]. This has some performance impact and it will be discussed later in the performance section 6.3.

2.8 Example

Fig. 5 illustrates a simple edge control system layout where camera is connected to CrossCore XM controller and video stream is displayed on CCpilot v700 display. Corresponding ROS2 application example is illustrated in Fig. 2. Camera Node reads images from camera sensor and publishes images to Image topic. Inference Node is the subscriber of the Image topic and runs inference on the received image. Then it publishes to Detection topic the detection result as an image with bounding boxes around detected objects. Display Node is the subscriber of Detection topic and it displays the detection results to the user.

3 Data Distribution Service

The greatest shortcomings regarding ROS1 industrial use was its middleware implementation. In ROS2 this has been considered from beginning. Instead of enhancing ROS1 implementation, a totally new middleware was selected for ROS2 from industry standard Data Distribution Service (DDS) [12].

DDS is proven technology with utilization in many mission-critical environments: [13]

- Battleships
- Large utility installations like dams
- Financial systems
- Space systems
- Flight systems
- Train switchboard systems

DDS uses Data-Centric Publish-Subscribe (DCPS) model. DCPS has "global data space" where the publisher writes the data and from where the subscriber reads the data [12]. This is analogous to ROS communication philosophy.

3.1 Wire Protocol specification

DDS standard does not specify how the actual communication is implemented. For this reason also Wire Protocol layer is needed to implement DDS and enable different vendors' DDS implementations work together. Wire protocol used with ROS2 is Real-Time Publish Subscribe (RTPS), a.k.a. DDS Interoperability Wire Protocol (DDSI-RTPS) [14]. RTPS has very similar working philosophy with DDS. Protocol definition was created by Object Management Group (OMG) [15].

3.2 ROS2 Middleware

ROS2 DDS/RTPS implementation is hidden behind ROS2 Middleware interface as illustrated in Fig. 1. ROS2 is agnostic to DDS and vice versa due interface abstraction. For this reason it is easy to plug in different vendor's DDS implementation to ROS2. The only requirement is that there exist ROS2 Middleware (RMW) implementation for binding ROS2 Middleware interface to DDS/RTPS implementation.

Table 1 lists currently available DDS/RTPS vendors that have RMW implementation. The default RMW implementation is eProsima's Fast RTPS.

Table 1. Supported RMW implementations [16].

Product name	License
eProsima Fast RTPS	LGPL
Eclipse Cyclone DDS	Eclipse Public License v2.0
RTI Connext	commercial, research
ADLINK Opensplice	Apache 2, commercial

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

4 Quality of Service

Quality of Service (QoS) policies define qualitative parameters for DDS communication. There are more parameters available on DDS standard than what ROS2 currently supports [17]. ROS2 QoS policies are categorized in this paper according to their effective domain as follows:

- Data persistence
- Communication robustness
- Real-time requirements
- Watchdogs

4.1 Data Persistence

There are three QoS policies that affect how well message data survives from node restarts and communication breaks.

Policy **History** defines how many old messages are kept in memory. Alternatives are *Keep last* and *Keep all*. If policy has value *Keep last*, then policy **Depth** determines the number of messages kept in memory. With *Keep all* all messages are kept in memory.

Another QoS policy that affect message data is **Durability**. It defines whether publisher is responsible for storing the sent messages on persistent memory or not. Possible values are *Transient local* which enables persistent memory saving and *Volatile* that prevents persistent storing of messages.

4.2 Communication Robustness

QoS policy **Reliability** defines the message transfer settings. Possible values are *Best effort* and *Reliable*.

With *Best effort* DDS/RTPS does not verify the transmission and it is analogous to UDP protocol.

With *Reliable* setting, the receiver is expected to acknowledge the received data. If no acknowledgement is received by sender, DDS/RTPS resends the message. *Reliable* is analogous to TCP/IP protocol.

4.3 Real-Time Requirements

It should be noted that in this paper term Real-Time (RT) refers to communication middleware's RT features, not to the RT application running on RT OS. ROS2 RT requirements ensure that the published messages reach the target node with given time constraints. And if that is not the case, both the publisher and the subscriber will be aware of that.

ROS2 RT features are controlled by QoS policies and there are currently two QoS policies that directly affect RT requirements.

Policy **Deadline** defines the maximum allowed time between subsequent messages. Both the message publisher and the subscriber will receive notification in case the message deadline is violated.

Policy **Lifespan** defines the time from message sending until it expires. If receiver does not receive the message before it expires, the message is silently dropped.

4.4 Watchdogs

Watchdog QoS policies help to identify if an individual ROS2 node is alive or not.

Policy **Lease Duration** determines the timeout after node's publisher is regarded as "dead", or in failure state, after receiving its last "liveliness" message.

Policy **Liveliness** determines how a node is determined to be in state "alive". Possible alternatives are *Automatic* and *Manual by topic*.

When alternative *Automatic* is selected and node publishes any message, then all the node's publishers are regarded as be in state "alive"

With alternative *Manual by topic* node's every publisher is responsible for its own state of being alive. When publisher sends a message, it is regarded as be in state "alive".

4.5 QoS Profiles

QoS policies can be grouped in QoS profiles. QoS Profiles identify common use cases and work as "shortcuts" to select suitable QoS policies. ROS2 has the following predefined QoS profiles [18]:

- Default profile – similar to ROS1 middleware behaviour with reliable message transfer and single last sent message kept in memory.
- Service profile – used with services. Reliable communication and no persistent message history to prevent outdated service requests being received after service server restart.
- Sensor profile – used with sensor nodes where it is important to transfer data quickly even some messages might be lost. *Best effort* reliability and small queue depth policies are used.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

- Parameter profile – used with parameters. Similar to Service profile, but with deeper queue length to catch parameter requests even if client has temporary network issues.

5 Security

The DDS Security Specification [19] introduces various security related plugins and policies which complement the "basic" DDS specification. ROS2 utilises these security additions in order to secure the communication between the nodes. Currently, only the mandatory parts of the specification are incorporated into the ROS2. This comprises the following plugins [20]:

- Authentication – To identify the communication participants.
- Access control – To restrict the communication participants' rights to certain resources in the ROS2 network.
- Cryptographic – To encrypt the messages, and perform any cryptography related operations.

The DDS Security is based on Public Key Infrastructure (PKI). X.509 certificates are used to bound the participants' public key to a specific name. Each x.509 certificate must have a signature chain to a specific Certificate Authority (CA) that the Authentication plugin is configured to trust. The access control is specified in XML-documents, called DDS governance and permissions files.

The implementation of the specification fulfilling security features is entirely on the ROS2 middleware providers' responsibility. Thus, there might be performance and security alterations between the RMW implementations. However, the comparison of the security aspects between different RMW implementations is out of the scope of this paper.

6 General evaluation

6.1 Documentation

There are public documentation for ROS2 in index and design sites. However, the content of the sites could be more comprehensive. Sometimes it is hard to find answers to specific questions from ROS2 official documentation. In some pages there are references that the work is still in progress and this is understandable. Maybe it changes when the ROS2 community grows and the demand for documentation increases as well.

One motivation for extended ROS2 background information in this paper, is to collect high-level basic knowledge of ROS2 in one place.

6.2 Design

Generally it is very good approach that the ROS2 middleware is not self-implemented, as was in ROS1, but an industry standard DDS is utilized. However, one downside about using this black-box is that it is more hard to find out answers to peculiarities in behaviour in comparison to situation middleware was open-source software where anybody could see from source code how it is implemented.

6.3 IDL data conversions

ROS2 and DDS messages are not directly compatible due different interfaces. For this reason an explicit conversion is needed between ROS2 and DDS interfaces. It should be noted that this conversion has to be done twice, once at the sending-end and once at the receiving-end. The performance impact of the conversion has been studied in [21] and it is illustrated in Fig 6.

The results confirm that with smaller messages, the conversion time is not significant, while with the larger messages it is significant. This result leads to a conclusion that ROS2 should preferably be used with smaller messages. Conversion delay might not be that significant in all applications and thus the larger message sizes could be used as well.

6.4 Flexible architecture

The interesting part with the ROS2 application is that there are no dependencies between the nodes. This means it is possible to modify the architecture and move individual nodes to different machine. If no hardware dependencies exist, the only modification to do is to load and run the ROS2 package on a new device and all the communication adapts to a new situation. This promotes genuine re-usability with ROS2 packages.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

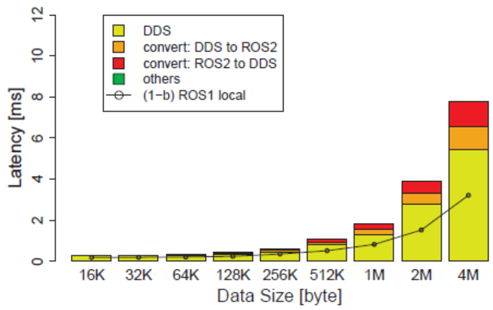


Fig. 6. ROS middleware performance comparison from [21] as a function of message size. Yellow is elapsed time with ROS2 DDS middleware, orange is the ROS2 message conversion time at the receiver, red is the ROS2 message conversion time at the sender, green is ROS2 other (insignificant) and line is ROS1 without DDS.

Table 2. Specification of test setup.

Nodes	Topic Publisher	Topic Subscriber
Machine	Dell Latitude E6520 (x86)	NVidia Jetson TX2 computer (ARM64)
Host OS	Linux Ubuntu 18.04 LTS	Linux Ubuntu 18.04 LTS
Memory	15GB	8GB
CPUs	Intel i7-2760QM	Denver 64-bit CPUs + Quad-Core A57 Complex
ROS2	Eloquent	Eloquent
DDS	eProsima Fast RTPS	eProsima Fast RTPS
Network swich	ZyXEL GS-1055 v2 Gigabit switch	ZyXEL GS-1055 v2 Gigabit switch
Network cable	Unshielded	Unshielded

7 Performance evaluation

ROS2 performance evaluation is divided into following sections:

- Data size impact on message transfer
- Data transfer latency
- ROS2 node performance
- CPU utilization

Table 3. Used QoS and security settings.

Configuration	QoS reliability	QoS history	Security
Unsecured reliable	Reliable	5000	No
Secured reliable	Reliable	5000	Yes
Unsecured best-effort	Best-effort	1	No
Secured best-effort	Best-effort	1	Yes

7.1 Data size impact on message transfer

The purpose of this evaluation is to find out how topic size and message sending frequency affect on the message throughput in practice. We are using test setup of single topic publisher and subscriber that are located on different physical computers and connected by Ethernet network. Table 2 describers the test SW and HW in more detail.

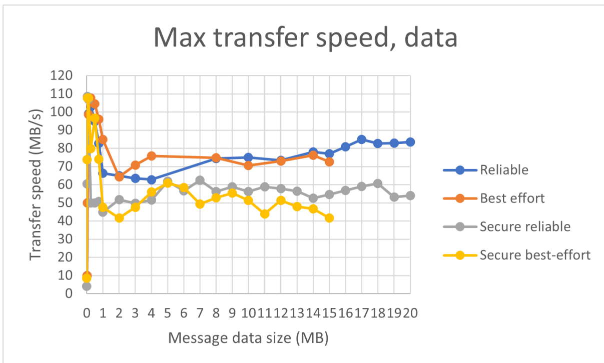


Fig. 7. Test setup maximum data transfer speed for all data size.

7.1.1 Data size affect on transfer speed

Data size tests are run using four different QoS and security setting groups as illustrated in Table 3.

It can be seen in Fig. 7 that the maximum transfer speed remains on the same level regardless of the messages size. However, with small message size (Fig. 9) the data rate is higher probably since data fits into a single IP message, while with larger sizes it is divided into several packages.

It can also be seen from the results that the security features require more data processing reducing the maximum transfer speed around 20MB/s.

One a bit surprising, but consistent observation with best-effort QoS reliability setting is that the topic's with message size exceeding 15MB could not be transmitted successfully at all. There is no clear answer yet to this behavior, but it might be that the message is divided into so many IP packages that none of complete topic messages could be sent successfully. Fig. 8 illus-

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

trates from one test run how large proportion of the messages were lost using different message sizes. Even there is lot of variance in the message loss % values, it can be seen that the trend is increasing while the message size increases, so loosing all the message data as message size increases is possible.

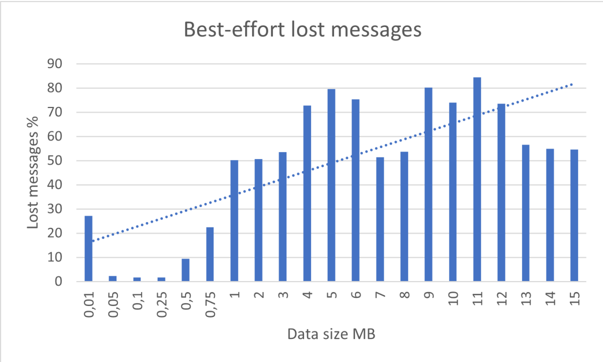


Fig. 8. Proportion messages lost during transfer from all the sent messages using Best-effort Reliability QoS setting.

For this reason, QoS reliability setting Best-effort does not seem feasible when dealing with large message data. It should be noted that we are using only the default DDS implementation, eProsima's Fast RTPS in these tests and it is possible that the similar effect is not occurring with other DDS implementations. However, that comparison is outside the scope of this paper.

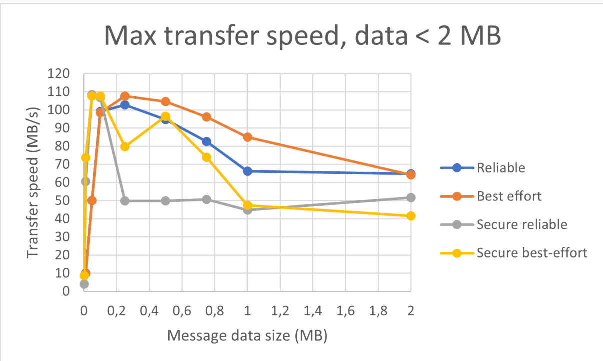


Fig. 9. Test setup maximum data transfer speed when data is smaller than 2 MB.

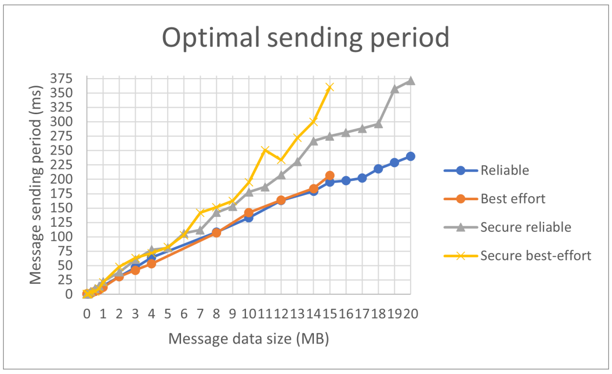


Fig. 10. Test setup optimal message sending period times for all data size.

7.1.2 Data size affect on choosing optimal message sending period

When considering the optimal message sending period time in this test, one tries to minimize the number of lost or re-sent messages thus obtaining the more optimal network bandwidth usage and overall performance. Fig. 10 illustrates the optimal sending rate for different QoS and security settings with all the measured message sizes.

It can be seen that with Reliable transfer, the message sending can be slightly faster than with Best-effort because it looses less messages which is a value factor when finding the optimal sending period in this test. However, it does not mean, Reliable setting would be faster than Best-effort, on the contrary, but Reliable is faster here simply since it looses less of the sent data. So in real application it depends on the task which reliability setting should be used and should the fast response time (higher speed) or most successfully transmitted data frames (reliability) be preferred over the other.

Security handling brings additional overhead of encryption and decryption as well as having larger encrypted messages sent over the network. These mean that the messages sent using secure transmit should be sent with smaller frequency than when sending messages without encryption in order not to waste the bandwidth.

7.2 Data transfer latency

Latency between message sending and receiving in topic based communication was examined using both secure and unsecure communication. The messages of different size were published to topic with exactly one publisher

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Table 4. Specification of secure communication delay test setup.

Nodes	Image Publisher	Image Subscriber
Machine	Dell Latitude E7470 (x86)	Dell Precision M4800 (x86)
Host OS	Windows 10	Windows 10
Guest OS	Ubuntu 18.04	Ubuntu 16.04
ROS2	Eloquent	Dashing
DDS	eProsima Fast RTPS	eProsima Fast RTPS
Memory (Guest)	16GB (6GB)	32GB (12GB)
Total CPUs	Intel i7-6600U 2.60GHz (4 cores)	Intel i7-4710MQ 2.50GHz (8 cores)
Guest CPUs	2.60GHz (2 cores)	2.50GHz (4 cores)
Network device	TP-LINK Archer C50 100 Mbps Router	
Network cable	Ethernet Cat 5e	

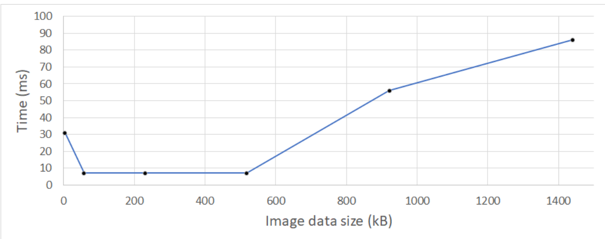


Fig. 11. Average delay added by secure communication in comparison to unsecure communication with different message size.

and one subscriber. Table 4 describes the used test setup and hardware characteristics.

The other computer was set to publish messages of type "sensor_msgs/Image" to a topic which the other subscribed to. Only "best effort" transmission policy was examined. Both nodes outputted their system time on image publish or arrival. The system time was synchronized using Network Time Protocol (NTP).

The results indicate that the overhead added by message encryption is not significant, although it starts to increase when the message size grows. Fig. 11 illustrates the results. It should be noted that the mentioned data size on the horizontal axis actually refers to the "sensor_msg/Image" message type's data part size and not actual UDP/IP message size which is somewhat larger.

7.3 C++ node performance

Single ROS2 C++ nodes execution was examined to gain knowledge of their performance and identify time consuming functionalities. The tests were conducted for simple publisher and subscriber nodes which were

Table 5. Specification of ROS2 node execution test setup.

Execution unit	Virtual Machine	Jetson Nano
Hardware	Dell Latitude E7470 (x86)	NVIDIA Jetson Nano System-on-Module (ARM64)
Host OS	Windows 10	Ubuntu 18.04
Guest OS	Ubuntu 18.04	-
DDS	eProsima Fast RTPS	eProsima Fast RTPS
Memory (Guest)	16GB (6GB)	4GB
Total CPUs	Intel i7-6600U 2.60GHz (4 cores)	ARM Cortex-A57 1.43GHz (4 cores)
Guest CPUs	2.60GHz (2 cores)	-

both communicating with only one topic. All the exchanged data was kept tiny by size, just 16 character long strings. The message type in ROS2 is standard string (std_msgs/msg/String).

All the communication was conducted on a single system, that is, publishers and subscribers were run on the same physical computer. Their QoS was set as reliable message transfer with 10 last sent messages kept in memory. As DDS implementation, the Fast RTPS was utilised. System time was logged using C++ standard library's chrono time library. The tests were executed on a virtual machine and Jetson Nano board, both running Ubuntu 18.04 Linux distribution and utilising Eloquent version of ROS2. Details of the test hardware is provided in table 5.

For the publisher test, a single subscriber was started and set to listen the common topic. The functionality provided by ROS2 client library (count_subscribers) was utilised by the publisher to ensure discovery of the subscriber. In addition, the publisher was forced to wait for one second before starting to poll the subscriber count. After subscriber was discovered a single message was sent to that topic. Then the publisher was terminated. The number of simultaneously launched publishers was altered and the tests were run in both unsecure and secure mode.

The publisher's execution time was measured after following code execution points:

– ROS2 client library init called
– node's construction finished
– subscriber found
– message published
– ROS2 client library shutdown called (total execution time)

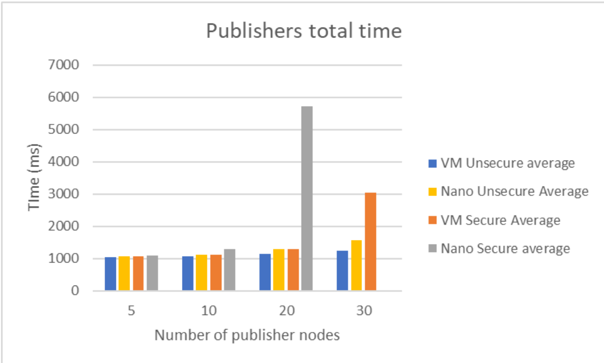


Fig. 12. Comparison of total execution time of publishers in milliseconds.

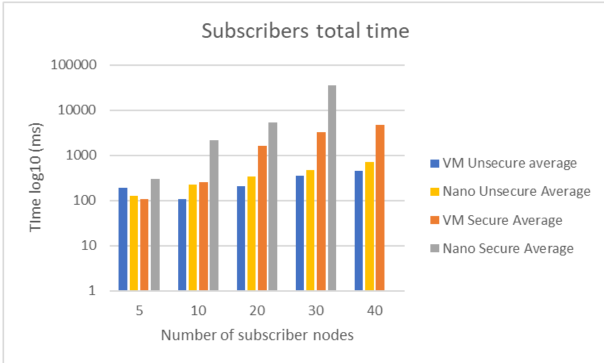


Fig. 13. Comparison of total execution time of subscribers in milliseconds. Please note the logarithmic scale.

Table 6 shows average times took by the publisher nodes to reach the specified checkpoints and Fig. 12 illustrates the total execution times. In the results, the intentional one second delay after node construction should be taken into account.

Results show that the time for constructing the publisher object slightly increases when more nodes are launched simultaneously. Most notable increase is however experienced during the subscriber search in secure mode. When the Nano has over 20, and the virtual machine over 30 publishers, the time rises drastically. Nano could not properly execute the 30 simultaneous publishers and thus has no corresponding records.

For the subscriber test, a single publisher was started and set to publish messages to the common topic at 20 Hz. The subscriber was started afterwards and set to listen the common topic. After receiving one message the subscriber was terminated. The number of simultaneously launched subscribers was altered and the tests were run in both unsecure and secure mode.

Quite the similar to the publisher test, single subscriber's execution time was measured after following code execution points:

– ROS2 client library init called
– subscribed to topic
– message received
– ROS2 client library shutdown called (total execution time)

Table 7 shows average times took by the subscriber nodes to reach the specified checkpoints and Fig. 13 illustrates the total execution times.

The results show surprisingly steep rise in message receiving times in secure mode on Jetson Nano. This is most probably indication that the Nano is incapable of handling such a high number of processes at once. On virtual machine, the communication delay behaves quite linearly. However, the difference between secure and unsecure communication is notably high when the number of subscribers increases. Being over 10 times slower is certainly alarming result and should be further investigated.

7.4 CPU utilization

In the last test setup we are measuring the CPU utilization by varying the number of simultaneous topic publishers.

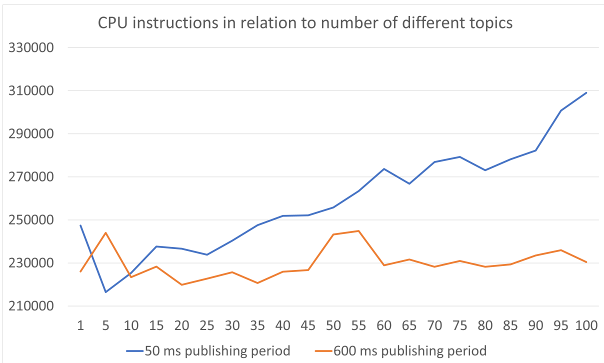


Fig. 14. Results from CPU utilization test with 50 ms and 600 ms publishing periods.

In Fig. 14 we can see that the CPU utilization is more or less linearly dependent on the number of publishers which is expected result. This gives us more confidence that there are no unlinear peculiarities in ROS2 C++ library implementation regarding the number of topic publishers.

In Fig. 14 there are measurements using two different publishing frequencies 50 ms and 600 ms. Results

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Table 6. Average checkpoint times in milliseconds for both devices, secure and unsecure communication and different number of publishers (N). Difference to previous in parentheses.

Device	Secure	N	Init	Node constructed	Subscriber found	Message published	Terminated (Total)
VM	false	5	14	27 (+13)	1028 (+1001)	1028 (+0)	1036 (+7)
Nano	false	5	20	57 (+37)	1058 (+1000)	1059 (+1)	1073 (+14)
VM	false	10	29	58 (+29)	1059 (+1001)	1059 (+0)	1071 (+12)
Nano	false	10	39	104 (+65)	1104 (+1000)	1104 (+0)	1125 (+21)
VM	false	20	61	123 (+62)	1124 (+1001)	1124 (+0)	1138 (+14)
Nano	false	20	103	253 (+150)	1253 (+1000)	1253 (+0)	1293 (+39)
VM	false	30	103	236 (+133)	1237 (+1001)	1237 (+0)	1245 (+8)
Nano	false	30	190	503 (+313)	1503 (+1000)	1503 (+0)	1563 (+60)
VM	true	5	16	50 (+33)	1050 (+1000)	1050 (+0)	1066 (+16)
Nano	true	5	24	66 (+42)	1067 (+1001)	1067 (+1)	1095 (+28)
VM	true	10	42	108 (+66)	1109 (+1001)	1109 (+0)	1126 (+17)
Nano	true	10	76	195 (+119)	1262 (+1067)	1262 (+0)	1299 (+37)
VM	true	20	94	242 (+147)	1260 (+1018)	1260 (+0)	1287 (+28)
Nano	true	20	125	322 (+198)	7523 (+7201)	7523 (+1)	7685 (+162)
VM	true	30	111	285 (+173)	3025 (+2740)	3025 (+0)	3046 (+21)

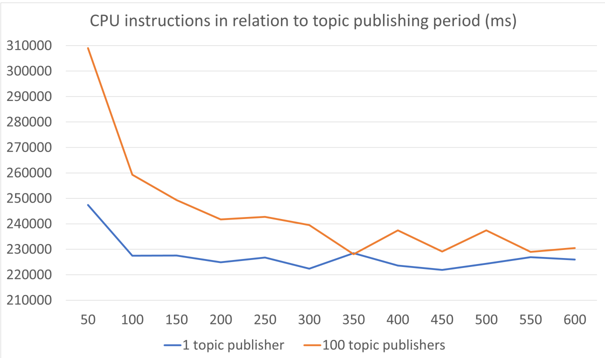


Fig. 15. Results from CPU utilization test having nodes with 1 and 100 topic publishers.

confirm that having fast publishing frequency causes higher CPU utilization which is also as expected.

Fig. 15 illustrates the situation where we keep the topic publisher count constant and only vary the publishing frequency. Here we see that the faster the topics are published, the greater impact is with the number of topic publishers. On the other hand, having lower frequency publishers reduces the difference in CPU utilization to insignificant.

8 Conclusions

ROS2 offers lots of tools for creating distributed computing systems that can be used in the edge environment. ROS2 isolates software components promoting code re-usability and cleaner SW architecture. ROS2 is an interesting candidate when choosing new platform for application development.

This paper focuses on explaining the basic structures of ROS2 application and evaluates the performance when using these basic structures. This will give user an overview of what ROS2 is all about and enable faster adoption of ROS2 working philosophy.

We did also made a practical evaluation of what kind of issues potentially rise when using ROS2 and default eProsima Fast-RTPS DDS implementation.

We have identified that the ROS2 contains necessary Real-time and security features that enable its use on edge devices. However, the security features should be used with care as they increase the communication and processing overhead significantly. It is best to choose only the necessary communication to be secure while keeping all the non-security-critical communication unencrypted. Also the number of publishers and subscribers should be limited when using security features as it might make the system unusable slow. Probably the best approach is to evaluate first the present hardware capabilities to find its limitations and only then make the final architectural choices how and where the ROS2 application nodes should be implemented.

Another interesting evaluation results have to do with the topic message size. When using Best-effort QoS reliability setting, we experiences total communication break after 15 MB message size. We also also experienced loss of large parts of the messages even before reaching this limit. This leads us to conclusion that with larger message sizes it is better to use Reliable QoS reliability setting.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Table 7. Average checkpoint times in milliseconds for both devices, secure and unsecure communication and different number of subscribers (N). Difference to previous in parentheses.

Device	Secure	N	Init	Subsc.	Message received	Terminated (Total)
VM	false	5	11	24 (+13)	182 (+158)	187 (+5)
VM	true	5	13	43 (+29)	99 (+56)	105 (+7)
Nano	false	5	20	46 (+26)	117 (+71)	125 (+8)
Nano	true	5	26	85 (+59)	285 (+201)	298 (+12)
VM	false	10	33	70 (+37)	99 (+29)	106 (+7)
VM	true	10	39	104 (+64)	238 (+134)	250 (+13)
Nano	false	10	41	96 (+54)	210 (+114)	221 (+11)
Nano	true	10	64	167 (+104)	2141 (+1973)	2187 (+47)
VM	false	20	74	152 (+78)	190 (+39)	203 (+13)
VM	true	20	80	214 (+134)	1609 (+1395)	1621 (+13)
Nano	false	20	119	267 (+148)	294 (+27)	339 (+45)
Nano	true	20	157	363 (+206)	5139 (+4776)	5207 (+68)
VM	false	30	123	257 (+134)	323 (+67)	346 (+23)
VM	true	30	124	286 (+162)	3198 (+2912)	3211 (+13)
Nano	false	30	174	370 (+196)	410 (+40)	462 (+52)
Nano	true	30	227	606 (+379)	34902 (+34297)	35084 (+181)
VM	false	40	169	357 (+188)	406 (+49)	452 (+46)
VM	true	40	145	366 (+221)	4626 (+4259)	4641 (+16)
Nano	false	40	250	557 (+307)	630 (+74)	723 (+92)

For the future ROS2 research, it is interesting to combine ROS2 with other technologies like Docker and Cloud computing. Both of these will enhance ROS2 use cases in real world scenarios by offering significant improvements in the edge-device deployment-strategies enabling ML model training data collection and Over-the-Air (OTA) updates for the ROS2 nodes and ML models. However, we need to do get similar performance benchmarking also for running ROS2 with the containers to see how much it brings penalty on resource consumption and startup times.

Another important future research topic is to compare network performance between clean office environment and in a noisy industrial environment where there are interfering factors that reduce the network throughput.

References

[1] Berger E, Wyrobek K. Stanford Personal Robotics Program. 2008.
URL http://personalrobotics.stanford.edu/

[2] Open Source Robotics Foundation, Inc. (OSRF). 2012.
URL https://www.openrobotics.org/

[3] Foote T. A History of ROS. 2020.
URL https://www.theconstructsim.com/history-ros/

[4] Scott K, ROSorg. ROS Metrics Report. 2020.
URL https://discourse.ros.org/t/2020-ros-metrics-report/16394

[5] ADACORSA. 2020.
URL https://adacorsa.eu/

[6] Quigley M, Gerkeyy B, Conleyy K, Fausty J, Footey T, Leibsz J, al et. ROS: an open-source Robot Operating System. In: *ICRA Workshop on Open Source Software*. 2009; .

[7] ROS2 Package Index. 2020.
URL https://index.ros.org/packages/

[8] Thomas D. A universal build tool. 2020.
URL http://design.ros2.org/articles/build_tool.html

[9] CrossControl. Products. 2020.
URL https://crosscontrol.com/en-us/products

[10] ROS interfaces. 2020.
URL https://index.ros.org/doc/ros2/Concepts/About-ROS-Interfaces/

[11] Mapping between ROS interface types and DDS IDL types. 2020.
URL https://design.ros2.org/articles/mapping_dds_types.html

[12] Data Distribution Service. *Tech. rep.*, Object Management Group.

[13] Woodall W. ROS on DDS. 2020.
URL http://design.ros2.org/articles/ros_on_dds.html

[14] DDS Interoperability Wire Protocol. *Tech. rep.*, Object Management Group. 2019.
URL https://www.omg.org/spec/DDSI-RTPS/2.3/PDF

[15] OMG. Object Management Group. 2020.
URL https://www.omg.org

[16] RMW implementations. 2020.
URL https://index.ros.org/doc/ros2/Concepts/DDS-and-ROS-middleware-implementations/

[17] About Quality of Service settings. 2020.
URL https://index.ros.org/doc/ros2/Concepts/About-Quality-of-Service-Settings/

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

[18] Fernandez E. ROS 2 Quality of Service policies. 2020. URL https://design.ros2.org/articles/qos.html
[19] DDS Security. *Tech. rep.*, Object Management Group. 2018. URL https://www.omg.org/spec/DDS-SECURITY/1.1
[20] Fazzari K. ROS 2 DDS-Security integration. 2021. URL https://design.ros2.org/articles/ros2_dds_security.html
[21] Maruyama Y, Kato S, Azumi T. Exploring the Performance of ROS2. In: *International Conference on Embedded Software (EMSOFT)*. 2016; . URL https://ieeexplore.ieee.org/document/7743223

This work is a part of ADACORSA project, that has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876019. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Netherlands, Austria, Sweden, Portugal, Italy, Finland, Turkey national Authorities.

ISBN: 13 978-952-5183-59-7

Tero Hietanen OAMK, Manne Tervaskanto OAMK, Joni Jämsä Centria, Outi Rask TAMK ja Niko Ristimäki SeAMK

Automaatiokoulutuksen sertifiointi

Tiivistelmä: Automation in Network (Aune) hankkeessa SeAMKin koordinoimana toteutettiin 14 ammattikorkeakoulun yhteinen hanke, jonka tarkoituksena on uudistaa ja tehostaa automaatiotekniikan opetusta ja opiskelua ammattikorkeakouluissa. /1/ Hankkeen luonnollisena jatkona on tuotetun koulutusaineiston pohjalta kehitetyn yhteisen koulutusmallin kehittäminen. Keskeistä nyt esitetyssä mallissa on yritysten tarpeista nouseva sertifioitu koulutus, jota voidaan tarjota osana jatkuvan oppimisen koulutustarjontaa. Koulutuksen toteutuksessa korostuu korona-ajan monimuotoinen koulutus eli etäopetus, virtualisoidut ohjelmistot sekä etäkäytettävät laboratoriolaitteistot. Hanketta kehittävät yhteistyössä Oulun AMK, Tampereen AMK, Ylivieskan Centria AMK sekä Seinäjoen AMK.

Asiasanat: automaatiotekniikan koulutus, etäopetus, sertifiointi

***Kirjoittajat:** Tero.Hietanen@oamk.fi, Manne.Tervaskanto@oamk.fi, Joni.Jamsa@centria.fi, Outi.Rask@tuni.fi ja Niko.Ristimaki@seamk.fi

1 Johdanto

Tekniikan alan koulutuksissa on menossa suuria muutoksia, joissa opiskelijat sijoittuvat suurempien ja laajempien kokonaisuuksien alle. Esim. Tampereen Yliopistossa automaatio on konetekniikan osa. Ammattikorkeakouluissa automaatiotekniikka hajaantuu sähkötekniikkaan, konetekniikkaan sekä rakentamistekniikkaan. Lähitulevaisuuden näkymissä sekä yliopistoissa, että ammattikorkeakouluissa on todennäköisesti hakukohteena tekniikka. Automaatiota on kaikkialla, mutta toisaalta ei missään määräävänä. Koulutukset joissa annetaan pätevyys johonkin toimintaan ovat tässä mielessä vahvemmassa asemassa.

Hankkeen taustalle selvitettiin yritysten ja organisaatioiden kiinnostusta automaatiokoulutuksen sertifiointia kohtaan. Hankkeelle saatiin nopealla kartoituksella noin 30 yrityksen ja organisaation tuki. Aluksi selvitettiin automaation osaamisen standardeja ja osaamisvaatimuksia. Lupaavimmat hankkeen kannalta olivat ISA (International Society of Automation) ja ECC (European Competence Center), havaittiin myös, että esim. turva-automaatio ja erityisesti tietoturvan osalta koulutus kansallisesti on suppeaa.

ISA on luonut automaation osaamistarpeille kattavan sertifioinnin, joka sisältää 7 pääkohtaa, joissa kussakin on 5-10 erilaista määriteltyä osaamista. ISAn mallissa sertifikaatti on voimassa 10 vuotta myöntämisestä. Pätevyys osoitetaan kokeella, ISA tarjoaa myös erikoistumiskoulutuksen sertifikaattia varten. ISA jakaa automaatio-osaamisen jatkuvaan säätöön, tuotantoautomaatioon, kehittyneeseen säätöön, turvallisuuteen ja luotettavuuteen, automaation IT:hen, tuotannon käynnissäpitoon sekä projektin hallintaan. /2/

Euroopassa lupaava osaamisen määrittely on IEC-standardin 61499 pohjalta toteutettu hajautetun automaation määrittely. /3/ Suomessa Kunnossapitoyhdistys Promaint ry järjestää kaksi kertaa vuodessa eurooppalaisen kunnossapitoyhdistysten kattojärjestön, EFNMS:n valvomia sertifiointitilaisuuksia. Tentin vaatimukset perustuvat SFS-EN 15628 Kunnossapitohenkilöstön pätevöinti -standardiin. /4/ Artikkelissa kuvataan hankkeistettu koulutuskokonaisuus, joka voisi olla ”sertifioitu automaatiokoulutus”.

2 Hankkeen tavoitteet

Hankkeen päätavoitteena on toteuttaa automaatioalan osaajapulaan nopeasti vastaava täydennyskoulutus työvoiman jatkuvaan oppimiseen ja osaamisen kehittämiseen. Kehittyvä teollisuus tarvitsee automaatioalan osaajia nopeasti ja osaamisen jatkuvaa kehittämistä. Oulun, Seinäjoen, Tampereen ja Centria ammattikorkeakoulujen yhteistyössä kehitetty digitaalisen koulutuksen osaaminen valjastetaan tässä hankkeessa työelämän tarpeisiin vastaaviksi, laadullisesti tunnustetuiksi sertifioiduiksi tai muutoin tunnustetuiksi koulutuskokonaisuuksiksi.

1. Tunnistetaan kansalliset ja alueelliset osaamistarpeet yritysten näkökulmasta.

Luodaan automaatioalan osaamisista ja sisällöistä suositus. Työssä hyödynnetään alan kansainvälisiä ja EU:n laajuisia standardeja sekä automaation henkilösertifikaatteja.

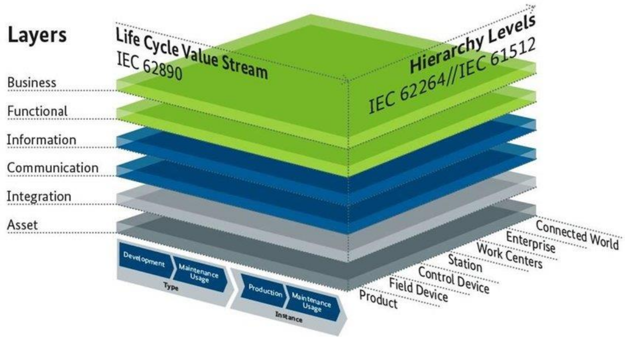
2. Työelämän osaajapulaan nopeasti vastaavat koulutuskokonaisuudet 60 op. Tavoitteena on toteuttaa sertifioitu automaatiokoulutus oppilaitosten yhteistyössä. Työnantajanäkökulmasta automaatiokoulutuksen sertifiointi tai muu laadunvarmistus on osaamisen laadun tae.
3. Pilotoidaan koulutuksia työttömillä sekä työelämässä jo olevilla. Tarvittaessa henkilöt voivat suorittaa osan kokonaisuudesta. Opinnot koostuvat pääosin eri oppilaitosten tarjoamista ammatillisista ja syventävistä etäopetuskokonaisuuksista. Koko koulutuksen suorittavien osalta 15op suoritetaan soveltavina työelämäopintoina.
4. Muodostetaan malli teollisuuden etujärjestöjen, alan organisaatioiden sekä sertifioinnista vastaavien tahojen kanssa henkilösertifioinnista osatavoite 1:n jaottelun pohjalta.
5. Kehitetään organisaatioiden osaamista ja tulevaisuuden koulutusmalleja sekä rakennetaan pysyvä oppilaitosyhteistyöverkosto.

3 Hankkeen toteutus

Hankkeen valmisteluvaiheessa tukensa antaneiden yritysten ja etujärjestöjen kanssa käydyissä keskusteluissa sekä toimijoiden osaamisen pohjalta muodostuivat koulutuksen sisällölle seuraavat osaamiskokonaisuudet:

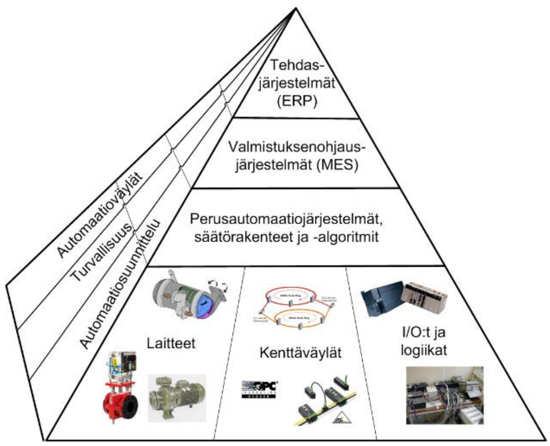
1. Prosessiautomaation suunnittelu ja projektointi
2. Tuotantoautomaation suunnittelu ja projektointi sekä robotiikka
3. Rakennusautomaation suunnittelu ja projektointi
4. Automaation elinkaaren hallinta ja käynnissäpito
5. Automaation informaatiotekniikka, tietojärjestelmät sekä IoT
6. Turva-automaatio sekä tietoturva

Koulutustarpeiden suuntausta kuvastaa RAMI-malli, joka kuvastaa automaatiohierarkiaa, elinkaaren hallintaa sekä digitaalisia kerroksia tiedon hallinnassa, kuva 1.



Kuva 1. Reference architecture model industry 4.0. /5/

Automaation perusteiden sertifiointi voisi yksinkertaistetummin pohjautua kuvan 2 mukaiseen jaotteluun. Sertifioidussa automaatiokoulutuksessa tulisi olla mukana aiemman lisäksi säätö- ja mittaustekniikan perusteet, sähkötekniikan ja elektroniikan perusteet, tietoliikennetekniikka ja väylät, prosessi-, kone- ja rakennustekniikan keskeisten yksikköprosessien perusteet.



Kuva 2. Automaatioalan koulutus muodostuu erilaisista perusteisiin liittyvistä kokonaisuuksista. Näiden perusteiden lisäksi on erilaisiin toimialoihin ja automaation sovelluskohteisiin liittyviä ominaisuuksia.

Opintojen toteutuksessa korostuvat etäopetusmateriaalit ja digitaaliset oppimisympäristöt. Perinteisempi opetus toteutetaan zoomissa tai vastaavassa sovelluksessa. Käytetyt ohjelmistot ovat virtualisoituja ja sovellussuunnittelun opiskelu onnistuu etätyönä. Laboratorioiden oppimisympäristöt ovat turvallisia etäkäyttää. Suuri osa arvioinnista perustuu interaktiivisiin oppimateriaaleihin.

Opintojen sisältöön kuuluu kirjallisia harjoitustöitä, ohjelmistoilla tehtäviä harjoituksia sekä projektitöitä. Opintoihin kuuluville yhteisille, työjärjestykseen merkityille osuuksille voi osallistua etäyhteyksien välityksellä. Opintojen toteutuksessa keskeisessä asemassa olevassa projektityössä suunnitellaan toteutuskelpoinen ratkaisu yrityslähtöiseen kehittämiskohteeseen, joka voi liittyä esimerkiksi robotiikkaan, automaattisiin tunnistustekniikoihin, virtuaalitekniikoihin tai IoT:iin. Aikaansaadun ratkaisun toimivuutta testataan ja esitellään mahdollisuuksien mukaan joko käytännössä demonstroiden, simuloiden tai muita havainnollistamistekniikoita hyödyntäen.

4 Yhteenveto

Hankkeen tuloksena on 60 opintopisteen laatuvarmistettu täydennyskoulutus automaatioalalle. Nopeavaikutteisella koulutuksella vastataan automaatioalan akuuttiin osaajapulaan ja parannetaan alan työllisyyttä. Alalla tunnistettu osaajapula kurotaan kiinni ajasta ja paikasta riippumattomalla koulutustarjonnalla, jolloin alalla työskentelevät voivat kehittää osaamistaan nopeasti.

Lyhytkestoisen koulutuksen laatu varmistetaan soveltamalla olemassa olevia, mutta ei käyttöön otettuja sertifiointikäytänteitä. ISA-mallin hyödyntäminen automaatio-osaamisen sertifioinnissa on uusi ja tärkeä avaus, jotta osaamisen taso on määriteltyä ja työmarkkinoilla on henkilösertifioituja automaatioammattilaisia. Hankkeen tuloksena on lisäksi ammattikorkeakoulujen henkilökunnan osaamistason nousu sekä pysyvä oppilaitosyhteistyö, jota pidetään yllä yhteisellä koulutustarjonnalla.

Runsaasta yritysten kiinnostuksesta huolimatta emme saaneet syksyllä 2020 jätetylle hankkeelle rahoitusta. Hankeideaa kehitetään edelleen ja konsortioon voi tulla mukaan eri koulutusasteita, yrityksiä ja yhteisöjä, jotka näkevät hankeidean kannatettavana. Jatkamme idean kehittämistä ja terävöittämistä seuraavissa soveltuvissa hauissa.

Liitteet

[1] https://www.seamk.fi/yrityksille/tki-projektit/automation-in-network/ 24.10.2020
[2] https://www.isa.org/training-and-certification/isa-certification/cap 24.10.2020
[3] http://daedalus.iec61499.eu/index.php/en/news/14-the-european-competence-network-for-iec-61499
[4] https://promaintlehti.fi/Lehtiarkisto/Promaint-4-2019
[5] https://www.isa.org/intech-home/2019/march-april/features/rami-4-0-reference-architectural-model-for-industr

3

Juha Hirvonen*

Opetuksen ja hanketyön linkittäminen teollisen internetin laboratorion avulla

Tiivistelmä: Tutkimus-, kehitys- ja innovaatiotoiminta (TKI-toiminta) on saamassa yhä keskeisempää roolia ammattikorkeakouluissa. Kehityshankkeiden kautta saadaan toki tehtyä yhteistyötä paikallisten teollisuusyritysten kanssa, ja parhaimmillaan ne voivat vaikuttaa positiivisesti siihen, että insinööriopinnot pysyvät ajanmukaisina ja vastaavat paikallisen elinkeinoelämän tarpeita. On kuitenkin riski, että TKI-toimintaa ja opetustyötä tekee suurelta osin eri henkilöstö, ja näin ne eriytyvät toisistaan eikä mahdollisia synergiaetuja saada. Siksi on tärkeää, että ammattikorkeakoululla on jokin strategia TKI-toiminnan ja opetuksen linkittämiseksi. Seinäjoen ammattikorkeakoulun teollisen internetin laboratorio palvelee juuri tätä tarkoitusta: sen avulla saadaan kehityshankkeiden tulokset myös opetuskäyttöön. Tämä artikkeli esittelee laboratorion sekä sen viimeaikaisen käytön niin hankkeissa kuin opetuksessakin sekä näiden kytkeytymisen toisiinsa.

Avainsanat: insinöörikoulutus, teollinen internet, opetuslaboratorio, TKI-toiminta

***Yhteyskirjoittaja: Juha Hirvonen:** Seinäjoen ammattikorkeakoulu, E-mail: juha.hirvonen@seamk.fi

1 Taustaa

Vuosina 2014–2015 toimeenpantu ammattikorkeakoulu-uudistus suhteutti opetus- ja kulttuuriministeriön ammattikorkeakouluille myöntämän perusrahoituksen niiden toiminnan tuloksellisuuteen, tehokkuuteen ja vaikuttavuuteen. Tutkimus-, kehittämis- ja innovaatiotoiminnan (TKI) painoa yhtenä kriteerinä on lisätty vuodesta 2014 [1]. Vuodesta 2021 lähtien TKI-toiminnan paino ammattikorkeakoulujen kokonaisrahoituksessa on 19 % [2]. Oman paineensa TKI-toiminnan lisäämiseen tuo lisäksi korkeakoulutuksen ja tutkimuksen 2030 -visio, jossa on asetettu tavoitteeksi nostaa Suomen tutkimus- ja kehittämisinvestoinnit 4 % bruttokansantuotteesta [3].

TKI-hankkeiden tekeminen mahdollistaa hyvän vuoropuhelun paikallisen yritysmaailman kanssa. Työ- ja elinkeinoministeriön vuonna 2019 julkaiseman raportin mukaan 17 % pk-yrityksistä tekee yhteistyötä korkeakoulun tai tutkimuslaitoksen kanssa [4]. Yleisin yhteistyötaho on ammattikorkeakoulu: niiden kanssa yhteistyötä tekee 12 % yrityksistä, kun vastaavat luvut yliopistojen ja tutkimuslaitosten kanssa ovat 7 % ja 4 % [4]. Suosituin yhteistyömuoto tosin on koulutus, jonka osuus ammattikorkeakouluyhteistyöstä on 54 %, mutta myös innovaatiotoiminta (25 %), pilotointi (12 %) ja laboratoriotilojen ja -palveluiden käyttö (10 %) on erottuvat tilastoissa [4].

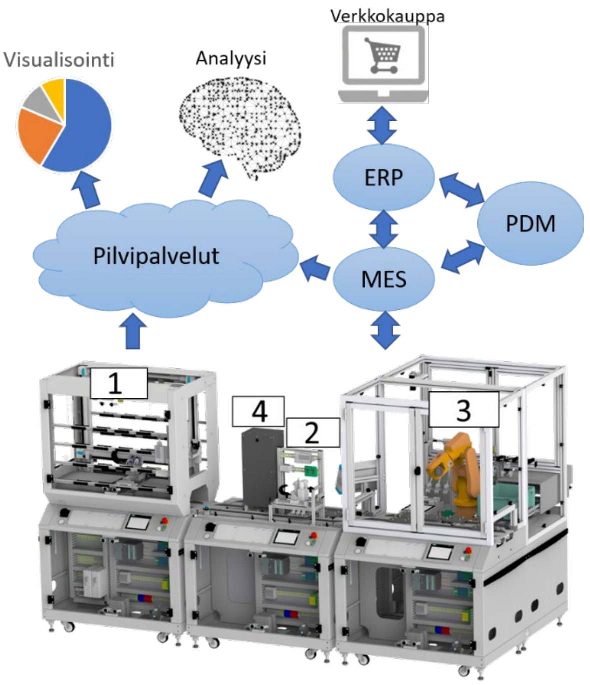
Parhaimmillaan TKI-toiminta ja yritysyhteistyö auttaa opetussisällön päivittämisessä vastaamaan alueen tarpeita. Sen kautta myös oppilaitoksen laboratorioiden laitekanta pysyy ajanmukaisena, sillä usein kehityshankkeet mahdollistavat laiteinvestointeja. Ongelmaksi voi kuitenkin muodostua se, että TKI-hankkeissa työskentelevät opetushenkilökunnan sijaan suurimmaksi osaksi erilliset TKI-asiantuntijat, ja näin hankkeista saatu uusi tieto siirtyy opetukseen hitaasti jos ollenkaan. Usein myös TKI-henkilöstön vaihtuvuus on suurempaa kuin opetushenkilöstön, joten hankkeissa saavutettu tieto on lisäksi vaarassa kadota ammattikorkeakoululta. Jotta hanketyön hedelmät olisivat paremmin opetuksen tukena, tulisi huolehtia opetushenkilökunnan ja TKI-henkilöstön riittävästä yhteistyöstä. Tämä voi tarkoittaa opettajien osallistumista hankkeisiin edes pienellä resurssilla tai vastavuoroisesti TKI-henkilöstön osallistumista opetukseen esimerkiksi vetämällä tiettyjä harjoituksia. Myös yleistä tiedonvaihtoa opettajien ja TKI-henkilöstön välillä tulisi tapahtua.

Tämä artikkeli esittelee Seinäjoen ammattikorkeakoulun teollisen internetin laboratorion käyttöä hanke- ja opetustyössä sekä näiden osa-alueiden yhteisiä rajapintoja. Kappaleessa 2 esitellään yleisesti teollisen internetin laboratorio sekä sen opetuskäyttö ja hankekäyttö. Kappaleesta selviää myös, kuinka hanketyön tulokset tukevat opetusta. Kappaleessa 3 kerrotaan esimerkkejä opiskelijoiden osallistumisesta itse hanketyöhön. Kappale 4 esittelee johtopäätökset.

2 Teollisen internetin laboratorio

Vuonna 2015 Seamkiin rakennettiin teollisen internetin laboratorio, jonka keskipisteenä on Feston miniatyrisoitu tuotantolinja (CP Factory). Tuotantolinja koostuu neljästä yksiköstä: korkeavarastosta, pora-asemasta, kokoonpanorobotista ja konenäkötarkastusasemasta. Tuotantolinjassa valmistetaan yksinkertaisia neljästä osasta (takakansi, painopiirilevy ja kaksi sulaketta) kostuvia matkapuhelinta muistuttavia tuotteita. Tuotantoa ohjataan itse tehdyllä tuotannonohjausjärjestelmällä (manufacturing execution system, MES), joka kommunikoi avoimen lähdekoodin toiminnanohjausjärjestelmän (enterprise resource planning, ERP) Odoon kanssa. On myös mahdollista käyttää valmistajan tarjoamaa MES-järjestelmää. Tuotetilauksia tehdään Odoon verkkokaupassa, josta tilaus välittyy ERP-järjestelmälle. Järjestelmänvalvoja voi hyväksyä tilauksia ja välittää ne edelleen MES-järjestelmälle, josta itse tuotanto voidaan käynnistää. Tuotevariantit eli eri kokoonpanot, jotka tuotantolinjalta on mahdollista tilata, voidaan määritellä joko verkkokaupan lisäosassa tai erillisessä tuotetiedonhallintajärjestelmässä (product data management, PDM), joka kommunikoi MES- ja ERP-järjestelmien kanssa.

Tuotantolinjassa on lisäksi erinäisiä antureita, joista kerätään dataa pilvipalveluun. Dataa kerätään myös MES-järjestelmästä. Kerättyä dataa voidaan visualisoida ja käyttää analyyseissa. Kuva 1 esittelee laboratorion tuotantolinjan ja siihen liitetyt tietojärjestelmäkytkökset. Laboratoriosta on pidempi esittely julkaisussa [5].



Kuva 1. Havainnekuva teollisen internetin laboratorion tuotantolinjasta tietojärjestelmäkytköksineen. Linjan yksiköt: (1) korkeavarasto, (2) pora-asema, (3) kokoonpanorobotti ja (4) konenäkötarkastusasema.

2.1 Laboratorion opetuskäyttö

Teollisen internetin laboratorion avulla opiskelijoille pyritään muodostamaan kokonaiskuva automaatiosta ja osoittamaan, että yksittäiset kurssit eivät ole erillisiä saarekkeita vaan osa suurempaa kokonaisuutta. Laboratoriota hyödynnetään laajasti automaatiotekniikan koulutusohjelman kursseissa. Tässä kappaleessa esitellään muutamia esimerkkejä niistä.

Ohjelmoitavan logiikoiden kurssilla tuotantolinja palvelee esimerkkinä automaatiojärjestelmästä. Sen avulla pystytään esittelemään opiskelijoille teollisuuden tietojärjestelmiä ja niiden toimintaa. Linjassa käytetään sekä Siemensin että Beckhoffin ohjelmoitavia logiikoita, joten se toimii hyvänä konkreettisena esimerkkinä.

Teollisen internetin kurssilla tietojärjestelmien toimintaa sekä datan keruuta esitellään vielä tarkemmin, ja niistä tehdään harjoituksia. Harjoituksissa opiskelija pääsevät tutustumaan eri pilvipalveluihin ja datan visualisointiin ja rakentamaan erinäisiä raportointinäkymiä.

Tuotannonohjauksen kurssilla laboratorion avulla käydään läpi CAD-ohjelman, tuotetiedonhallinnan ja toiminnanohjausjärjestelmän yhteispeliä. Opiskelijat tekevät itse valmistettavan tuotteen rakenteen eli määrittelevät komponentit, mistä se koostuu, tallettavat sen Teamcenter-tuotetiedonhallintajärjestelmään ja ajavat tilauksen tuotantolinjalle.

Kurssilla *konenäkömenetelmät- ja sovellukset* opiskelijat toteuttavat itse laboratorion suorittamat konenäkötarkastukset. Kokoonpanoyksikössä tarkistetaan kappaleen orientaatio valopöydällä ja konenäkötarkistuspisteellä tuotteen sisältämät komponentit kirkaskenttävalaistuksella. Opiskelijat saavat siis tyyppiesimerkit kahden eri valaistustekniikan käytöstä.

Projektityökurssilla opiskelijat tekevät ohjelmistoprojekteja laboratorion avulla. Esimerkkejä näistä ovat erilaiset web-pohjaisia käyttöliittymät ja mobiilisovellukset.

2.2 Laboratorion hankekäyttö

Hanketyössä laboratorio toimii testipenkkinä erilaisille tuotantoon liitettäville tietojärjestelmille ja teollisen internetin sovelluksille. Etelä-Pohjanmaalla on paljon valmistavan teollisuuden yrityksiä, joten oma kokonainen tuotantolinja – vaikka sitten miniatyrisoitu – tuo hankekumppanina uskottavuutta, sillä sen avulla voidaan demonstroida erilaisia tuotannossa sovellettavia tietoteknisiä ratkaisuita. Myös opetuksen kehittämiseen liittyvissä hankkeissa laboratorio on hyvä testipenkki. Tämä kappale antaa joitain esimerkkejä hankkeista, joissa linjaa on sovellettu.

Hankkeessa *Näkymätön näkyväksi* osallistuneille pk-yrityksille on laboratorion avulla esitelty toiminnanohjausjärjestelmän eri ominaisuuksia. Hankkeessa on myös kehitetty eteenpäin laboratorion omatekoista

ISBN: 13 978-952-5183-59-7

tuotannonohjausjärjestelmää ja esitelty kehitystyötä sekä järjestelmän käyttöä yrityksille. Lisäksi hankkeessa on demonstroitu lisätyn todellisuuden (augmented reality, AR) hyödyntämistä tuotannon tulosmittareihin: linjaan kiinnitettyjen AR-markkereiden avulla älypuhelimella voi katsella kuvaajien muodossa tuotannon tunnuslukuja kuten laitteen kokonaistehokkuuden tai yksittäisten työasemien työaikatietoja.

Hankkeessa *Enterprise digital twin platform* linjasta on rakennettu digitaalinen kaksonen. Linjan simulaatiomalliin on siis tehty rajapinnat myös MES- ja ERP-järjestelmille. Ideana hankkeessa on kehittää tapa erilaisten MES- ja ERP-järjestelmien testaamiseen simulaatiomallien avulla. Simulaatiomalli on rakennettu Tecnomatix Plant Simulation -ohjelmalla ja tarvittavat rajapinnat on ohjelmoitu Python-ohjelmointikielellä. Rakennettua digitaalista kaksosta ja sen toimintaa esitellään tarkemmin julkaisussa [6].

Hankkeessa *IoT Compass Hubin startti* linjaa on hyödynnetty alueen jatkuvan oppimisen kehittämiseen. Hankkeessa on työstetty mallia, jossa yritysten työntekijät voivat tulla laboratorioon opettelemaan tuotannon tietojärjestelmien sekä teollisen internetin sovellusten toimintaa joko keskenään tai yhdessä opiskelijoiden kanssa. Tällä tavoin saadaan myös tutustutettua opiskelijoita ja yritysten edustajia toisiinsa. Mallista on yksityiskohtaisempi selvitys julkaisussa [5].

3 Opiskelijoiden osallistuminen hanketyöhön laboratoriossa

Kuten kappale 2 osoittaa, laboratorion opetuskäytössä hyödynnetään jo itsessään runsaasti hanketyön tuloksia: laboratorion tietojärjestelmien kommunikointi toistensa kanssa on toteutettu hanketyönä, ja opiskelijat pääsevät hyötymään näistä tuloksista yhteensä kolmella eri kurssilla. Myös erilaisten tulosmittareiden ja raportointinäkymien käyttö ja hyödyntäminen on alkujaan hankkeissa hankittua osaamista, ja se on sieltä siirtynyt myös osaksi opetusta. Kun opetushenkilöstö on mukana hankkeissa ja tekee yhteistyötä TKI-henkilöstön kanssa, hankkeiden tulokset saadaan siirrettyä sopiville kursseille parantamaan opetustarjontaa.

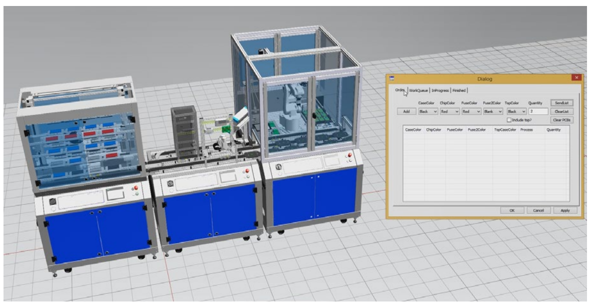
Joissain tapauksissa opiskelijat ovat olleet myös mukana TKI-toiminnassa laboratorion avulla. Tässä kappaleessa kerrotaan siitä kaksi esimerkkiä. Tulevaisuudessa tällaista TKI-perustaista oppimista tullaan hyödyntämään aiempaa enemmän [7].

3.1 Tuotannonohjausjärjestelmän ohjelmointi

Laboratoriossa käytettävän itse tehdyn tuotannonohjausjärjestelmän ensimmäinen versio on tehty opiskelijatyönä. Opiskelijat siis osallistuivat projektissaan hankkeeseen, jossa kehitettiin teollisen internetin opetusta. Projektinsa yhteydessä he selvittivät laitteiden välisen kommunikaation, suunnittelivat tuotannonohjausjärjestelmän ohjelma-arkkitehtuurin ja toteuttivat ohjelman eri osaset Python-ohjelmointikielellä. Python oli opiskelijoille uusi ohjelmointikieli, mutta käytännöllinen ja innostava aihe auttoi heitä omaksumaan sen hyvin nopeasti. Projektissaan he lisäksi oppivat työskentelemään tiiminä, jakamaan kehitysvastuuta ja käyttämään versionhallintaa sekä saivat perehtyä hyvin syvällisesti tuotannonohjauksen logiikkaan. Tuotannonohjausjärjestelmä jäi tosin osin kesken, mutta se viimeisteltiin vastavalmistuneen opiskelijan toimesta osana toista hanketta kuten kappaleessa 2.2 kerrottiin. Osia alkuperäisestä ohjelmasta kuten TCP/IP-sokettikommunikaation hoitavaa luokkaa on käytetty hyödyksi muussakin hanketyössä esimerkiksi viestinvälityksessä yhteistyöroboteille. Koska tuotannonohjausjärjestelmä on rakenteeltaan modulaarinen, sitä voidaan hyödyntää myös tietotekniikan opetuksessa siten, että ohjelmakoodista poistetaan yksittäinen osa, joka opiskelijoiden sitten pitää itse toteuttaa. Ohjelmointityöt, joissa laiterajapinta on mukana, ovat tunnetusti opiskelijoille mieluisampia niiden käytännönläheisyyden vuoksi. Tällaisia harjoitustöitä on tarkoitus teettää tulevaisuudessa.

3.2 Digitaalisten kaksosten rakentaminen

Opiskelijat, jotka ovat jääneet ilman harjoittelupaikkaa, ovat voineet suorittaa harjoittelunsa teollisen internetin laboratoriossa jonkin hankkeiden ja yritysyhteistyön kannalta oleellisen aiheen parissa. Laboratoriossa on myös muutamia tietokoneita, ja niissä on Siemensin simulaatio-ohjelmistot. Opiskelijavoimin on mm. tehty tuotantolinjan digitaalinen kaksonen 3D-työkaluilla työkalujen demonstroimiseksi, joka esitellään Kuvassa 2. Tällä digitaalisella kaksosella on ollut oma käyttöliittymänsä, jonka kautta tilauksia on tehty eli sillä ei ole ollut rajapintaa erilliseen ERP- tai MES-järjestelmään. Myös datan keräämistä digitaaliselta kaksoselta on kokeiltu opiskelijatyönä. Koska laboratoriossa vierailee



Kuva 2. Opiskelijan laboratorion tuotantolinjasta rakentama digitaalinen kaksonen.

usein yritysten edustajia, siellä työskentelevät opiskelijat pääsevät näyttämään osaamistaan heille ja parantavat mahdollisuuksiaan työllistyä tai saada harjoittelu- tai opinnäytetyöpaikka yrityksessä. Kerran opiskelija sai tällaisen vierailun yhteydessä kutsun työhaastatteluun. Ylipäätään laboratoriosta tehtyjä digitaalisia kaksosia hyödynnetään myös opetuksessa.

4 Johtopäätökset

TKI-toiminnan mukana ammattikorkeakoulun laboratoriot kehittyvät ja laitekanta kasvaa. Kun TKI-asiantuntijat ja opettajat työskentelevät yhdessä hankkeissa tai muuten pitävät säännöllistä yhteyttä, hanketyön tuloksia pystytään hyödyntämään opetuksessa. Näin saadaan yleensä lisää käytännön harjoituksia, harjoitustyöaiheita, projektityöympäristöjä sekä mahdollisia harjoitteluympäristöjä. Yleisesti insinööriopiskelijat pitävät käytännöllisestä opiskelusta ja laitteiden kanssa työskentelystä, joten hanketyöllä saavutetuilla tuloksilla voidaan vaikuttaa positiivisesti myös opiskelijatyytyväisyyteen. Tässä artikkelissa esiteltiin yksittäisen laboratorion käyttöä sekä osana opetusta että hanketyötä ja näiden kytkeytymistä toisiinsa useiden esimerkkien avulla.

Hankkeissa syntyneiden tulosten hyödyntäminen vaikkapa harjoituksena tai harjoitustyön aiheena on huomattavasti suoraviivaisempaa kuin opiskelijoiden integroiminen suoraan hanketyöhön. Siitäkin on muutamia kokemuksia, jotka esiteltiin artikkelissa, ja tätä tullaan tekemään tulevaisuudessa lisää [7]. Opiskelijoiden käytöstä osana hankkeita on kuitenkin huomioitava, että teetettävät tehtävät eivät ole liian kriittisiä hankkeen kannalta ja että opiskelijoille on varmistettu riittävä tuki. Parhaiten hankkeesta irrotetut tehtävät soveltuvat loppuvuosien projektitöiksi. Opiskelijatyöhön liittyvää riskiä hankkeelle voidaan pienentää sillä, että useampi ryhmä tekee oman toteutuksensa aiheesta. Näin onnistumisen todennäköisyyttä voidaan kasvattaa.

Mitä suuremman roolin hankkeista saatavat projekti- ja harjoitustyöt saavat opetuksessa, sitä merkittävämmäksi tulee ohjaus. Jos harjoitustöiden aiheet tulevat kulloinkin käynnissä olevista hankkeista, kurssin eri toteutuskertojen välille tulee helposti eroa. Opettajan pitää siis joka toteutuksen yhteydessä miettiä, mitä osaamistavoitteita harjoitustyön tekeminen tukee, ja mitkä osaamistavoitteet jäävät muuten opetettavaksi. Näin voidaan varmistua siitä, että joka vuosi käsitellään kuitenkin samat asiat – joskin eri tavalla.

Kiitokset

Tämä artikkeli on kirjoitettu osana opetus- ja kulttuuriministeriön rahoittamaa hanketta More start-ups and growth through digitalisation and artificial intelligence.

Lähteet

[1] Laakso, A., Stormi, I., Mikkonen, A., Ammattikorkeakoulujen TKI-toiminta Suomen kilpailukyvyn edistäjänä, Tieteessä tapahtuu, 2020, 38, 33–39

[2] Opetus- ja kulttuuriministeriö, Ammattikorkeakoulujen rahoitusmalli vuodesta 2021 alkaen, saatavilla: https://minedu.fi/documents/1410845/4392480/AMK_rahoitusmalli_fi/65aedf23-83a0-64f7-aed5-ebdf07bc9efe/AMK_rahoitusmalli_fi.pdf, viitattu 10.2.2021

[3] Opetus- ja kulttuuriministeriö, Visio 2030 työryhmien raportit, saatavilla: https://minedu.fi/documents/1410845/12021888/Visiotyo%CC%88ryhmien+yhteinen+taustaraportti_v2.pdf/d69fc279-d6a9-626d-deac-712662738972/Visiotyo%CC%88ryhmien+yhteinen+taustaraportti_v2.pdf, viitattu 10.2.2021

[4] Huovinen, P., Kärpänoja J., Pk-yritysten yhteistyö korkeakoulujen ja tutkimuslaitosten kanssa, Husso, K., (toim.), Työ- ja elinkeinoministeriö, Helsinki, 2019

[5] Lanz, M., Lobov, A., Katajisto, K., Mäkelä, P., A concept and local implementation for industry-academy collaboration and life-long learning, Procedia Manuf., 2018, 23, 189-194, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2018.04.015

[6] Hirvonen, J. Hakalahti, H., Ylihärsilä, M., Testing ERP and MES with Digital Twins, teoksessa: Proceedings of The 61st SIMS Conference on Simulation and Modelling, Virtual Conference (22–24 September, 2020, Finland), SIMS 2020, Finland, 2020, 379–384.

[7] Junell, P., Viljamaa, A., Nurmi, C., Mikkola, M., TKI-integroitu oppimiskonsepti digitalisaation, tekoälyn ja kasvuyrittäjyyden osaamisen edistämiseksi, teoksessa: Junell P., Hirvonen J., Sivula A., Rasku H., Saarikoski S. (toim.), SeAMK Tekniikan tutkimus, kehittäminen ja opetus rakentamassa alueellista innovaatioekosysteemiä, Seinäjoen ammattikorkeakoulu, Seinäjoki, 2020, 169–186.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Konsta Huhtala* and Lassi Paunonen

Output Regulation of Thermal Fluid Flows

Abstract: We consider temperature output tracking for a room model. Dynamics of the fluid within the room are governed by the Boussinesq equations, which we consider linearized around a steady state solution. We show that the system of PDEs complemented with a weighted average temperature observation form a regular linear system, thus a particular error-feedback controller achieves output tracking of sinusoidal reference signals. The theoretical results are illustrated through numerical simulations.

Keywords: distributed parameter systems, output regulation, fluid flow systems

***Corresponding Author: Konsta Huhtala:** Faculty of Information Technology and Communication Sciences, Mathematics and Statistics, Tampere University, PO. Box 692, 33101 Tampere, Finland, E-mail: konsta.huhtala@tuni.fi
Lassi Paunonen: Faculty of Information Technology and Communication Sciences, Mathematics and Statistics, Tampere University, PO. Box 692, 33101 Tampere, Finland, E-mail: lassi.paunonen@tuni.fi

1 Introduction

In this paper, we study a temperature output regulation problem for a two-dimensional room model with the room geometry depicted in Figure 1. Our control goal is to have a weighted temperature average over a certain subdomain Ω_y of Ω converge to a desired reference trajectory of the form

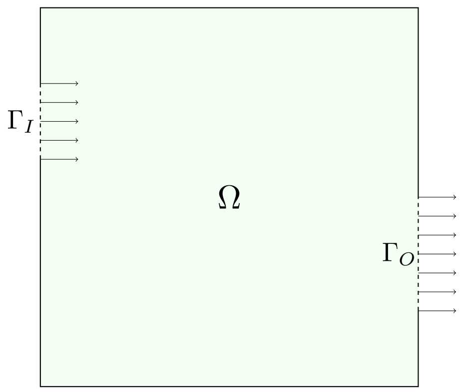


Fig. 1. Outline of the room

$$y_r(t) = a_0 + \sum_{k=1}^{q} (a_k \cos(\omega_k t) + b_k \sin(\omega_k t)), \qquad (1)$$

where for each $k = 0, 1, \dots q$, a_k and b_k are possibly unknown coefficients and $0 = \omega_0 < \omega_1 < \dots < \omega_q$ are known frequencies.

Fluid dynamics within the room are governed by the Boussinesq equations. While the stabilization problem for the Boussinesq equations has been studied, see [2, 11], results concerning output regulation are limited, see [1]. In this work, our approach to temperature output regulation is based on considering a linearization of the Boussinesq equations. We then achieve output tracking using a combination of control inputs acting on either the velocity or the temperature component of the system state, located either within the domain Ω or within the domain's boundary Γ. An alternative approach to output regulation based on the assumption of stationary velocity field has been considered in [6].

As the main result of this paper, we show that the room model is a *regular linear system* in the sense of [14]. This result implies that we have readily available controller designs, see e.g. [8], for robust temperature output regulation for the room model. We implement one such controller, called the low-gain controller, and illustrate its output tracking performance with a numerical example.

The paper is organized as follows. In Section 2, we present the considered room model and verify that the model forms a regular linear system. In Section 3, we first present the controller used in this paper to achieve temperature output regulation and then discuss alternative controller options. Finally, Section 4 contains a numerical example of temperature output tracking and is followed by conclusion of the paper in Section 5.

From now on, we denote by $\mathcal{L}(X, Y)$ the set of bounded linear operators from a Hilbert space X to a Hilbert space Y and denote by $D(A)$ the domain of a linear operator A.

2 The Room Model

The Boussinesq equations on the domain Ω depicted in Figure 1 are given by

$$\frac{\partial w}{\partial t} = \frac{1}{Re}\Delta w - (w\cdot\nabla)w - \nabla q + \hat{e}_2\frac{Gr}{Re^2}T + f_{bw}, \quad (2a)$$
$$0 = \nabla\cdot w, \quad (2b)$$
$$\frac{\partial T}{\partial t} = \frac{1}{RePr}\Delta T - w\cdot\nabla T + f_{bT}, \quad (2c)$$

where $w(\xi,t)$ is the fluid velocity, $q(\xi,t)$ is the fluid pressure, $T(\xi,t)$ is the fluid temperature, Re is the Reynold's number, Gr is the Grashof number, Pr is the Prandtl number, $\hat{e}_2 = [0,\ 1]^T$ indicates direction of the gravitational force and $f_{bw}(\xi)$, $f_{bT}(\xi)$ are an external body force and a heat source, respectively.

We define the following spaces to formulate and analyze the room model.

$$\begin{aligned} X_v &= \{v\in(L^2(\Omega))^2 \mid \nabla\cdot v = 0,\ (v\cdot n)|_{\Gamma_W} = 0\},\\ X &= X_v\times L^2(\Omega),\\ H_v &= \{v\in(H^1(\Omega))^2 \mid \nabla\cdot v = 0,\ v|_{\Gamma_W} = 0\},\\ H_\theta &= \{\theta\in H^1(\Omega) \mid \theta|_{\Gamma_W} = 0\},\\ H &= H_v\times H_\theta,\end{aligned}$$

where n is the unit outward normal of Γ and $\Gamma_W = \Gamma\setminus(\Gamma_I\cup\Gamma_O)$. Additionally, we assume that the inlet $\Gamma_I\subset\Gamma$ and the outlet $\Gamma_O\subset\Gamma$ are disjoint.

2.1 Problem Formulation

We assume that there exists a steady state solution

$$[w_{ss}(\xi), q_{ss}(\xi), T_{ss}(\xi)]^T \in H_v\times L^2(\Omega)\times H_\theta$$

to the Boussinesq equations (2) and linearize the equations around the steady state solution. The linearized translated system with control inputs $u_v(t)\in\mathbb{R}$ and $u_\theta(t)\in\mathbb{R}$, and control shape functions $b_v\in X_v$ and $b_\theta\in L^2(\Omega)$ is given by

$$\begin{aligned}\frac{\partial v}{\partial t} &= \frac{1}{Re}\Delta v - (v\cdot\nabla)w_{ss} - (w_{ss}\cdot\nabla)v - \nabla p\\ &\quad + \hat{e}_2\frac{Gr}{Re^2}\theta + b_v u_v, \quad v(\xi,0) = v_0(\xi)\in X_v, \quad (3a)\end{aligned}$$
$$0 = \nabla\cdot v, \quad (3b)$$
$$\frac{\partial\theta}{\partial t} = \frac{1}{RePr}\Delta\theta - w_{ss}\cdot\nabla\theta - v\cdot\nabla T_{ss} + b_\theta u_\theta, \quad (3c)$$
$$\theta(\xi,0) = \theta_0(\xi)\in L^2(\Omega), \quad (3d)$$

where $v(\xi,t) = w(\xi,t) - w_{ss}(\xi)$, $\theta(\xi,t) = T(\xi,t) - T_{ss}(\xi)$ and $p(\xi,t) = q(\xi,t) - q_{ss}(\xi)$. The boundary conditions related to (3a)-(3d) are

$$(\mathcal{T}(v,p)\cdot n)|_{\Gamma_I} = b_{v\Gamma}u_{v\Gamma}, \quad (3e)$$
$$(\mathcal{T}(v,p)\cdot n)|_{\Gamma_O} = 0, \qquad v|_{\Gamma_W} = 0, \quad (3f)$$
$$\frac{1}{RePr}\frac{\partial\theta}{\partial n}\Big|_{\Gamma_I} = b_{\theta\Gamma}u_{\theta\Gamma}, \quad (3g)$$
$$\frac{1}{RePr}\frac{\partial\theta}{\partial n}\Big|_{\Gamma_O} = 0, \qquad \theta|_{\Gamma_W} = 0, \quad (3h)$$

where $\mathcal{T}(v,p)$ is the fluid Cauchy stress tensor, $u_{v\Gamma}(t)\in\mathbb{R}$ and $u_{\theta\Gamma}(t)\in\mathbb{R}$ are control inputs, and $b_{v\Gamma}\in(L^2(\Gamma_I))^2$ and $b_{\theta\Gamma}\in L^2(\Gamma_I)$ are control shape functions.

We consider a temperature observation of the form

$$y(t) = \int_{\Omega_y} c_y(\xi)\theta(\xi,t)d\Omega_y, \quad (4)$$

where $c_y\in L^2(\Omega_y)$. The output regulation goal is for the observation to converge exponentially to the reference signal (1), i.e. there should exist constants $M_c, \omega_c > 0$ such that

$$\|y(t) - y_r(t)\| \le M_c e^{-\omega_c t}. \quad (5)$$

The exponential convergence should be achieved for any initial states v_0 and θ_0 and any reference signals of the form (1), but the constant M_c may depend on the initial states and the coefficients a_k and b_k in (1).

2.2 Model Analysis

Linearization of the Boussinesq equations has been used for stabilizing controller designs, and we get the following result directly from the stabilization focused work in [2].

Theorem 1. *The linearized translated Boussinesq equations* (3) *with the observation* (4) *can be formulated as an abstract linear system*

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0\in X, \quad (6a)$$
$$y(t) = Cx(t) \quad (6b)$$

with the input space $U = \mathbb{R}^4$ and the output space $Y = \mathbb{R}$, where $x = [v,\ \theta]^T$, $u = [u_v,\ u_\theta,\ u_{v\Gamma},\ u_{\theta\Gamma}]^T$, the operator A defined by

$$A\begin{bmatrix} v\\ \theta\end{bmatrix} = \begin{bmatrix}\mathbb{P}\left(\frac{1}{Re}\Delta v - (w_{ss}\cdot\nabla)v - (v\cdot\nabla)w_{ss} + \frac{Gr}{Re^2}\hat{e}_2\theta\right)\\ \frac{1}{RePr}\Delta\theta - w_{ss}\cdot\nabla\theta - v\cdot\nabla T_{ss}\end{bmatrix}$$

is the generator of a strongly continuous semigroup $\mathbb{T}_A$ on X with $\mathbb{P}$ denoting the Leray projector, c.f. [7, Lemma 2.2],

$$B = [B_\Omega \quad B_\Gamma], \qquad B_\Omega = [b_v \quad b_\theta]\in\mathcal{L}(\mathbb{R}^2, X),$$
$$C = \langle[0,\ c_y]^T, (\cdot)\rangle_{L^2(\Omega_y)}\in\mathcal{L}(X,Y).$$

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

The precise definitions of B_Γ and $D(A)$ can be found in [2]. We choose to omit the definitions here, since they would require introducing an extensive amount of notation. However, we note that the adjoint operator of B_Γ has the presentation

$$B_\Gamma^* \begin{bmatrix} v \\ \theta \end{bmatrix} = \begin{bmatrix} \langle b_{v\Gamma}, v\rangle_{L^2(\Gamma_I)} \\ \langle b_{\theta\Gamma}, \theta\rangle_{L^2(\Gamma_I)} \end{bmatrix} \qquad \forall \begin{bmatrix} v \\ \theta \end{bmatrix} \in D(A). \tag{7}$$

We now consider a further result for the abstract system formulation (6).

Theorem 2. *The system* (6) *is a regular linear system in the sense of [14].*

Proof. The observation operator C is bounded. As such, it suffices to show that the control operator B is admissible for $\mathbb{T}_A$, since [13, Prop. 4.4.6] then immediately implies existence of the limit

$$\lim_{s\in\mathbb{R}, s\to\infty} C(sI - A)^{-1}Bu$$

for any $u \in U$ and thus regularity of the system (6). From [11] we have that for a large enough $\lambda > 0$, fractional powers of the operator $\lambda I - A$ and its adjoint $\lambda I - A^*$ exist and

$$D((\lambda I - A)^{1/2}) = D((\lambda I - A^*)^{1/2}) = H.$$

Furthermore, by [2] we have the decomposition $A = A_2 + A_{10}$, where

$$A_2 = \begin{bmatrix} \frac{1}{Re}\mathbb{P}\Delta & 0 \\ 0 & \frac{1}{RePr}\Delta \end{bmatrix}$$

is self-adjoint and strictly negative, and for every $\begin{bmatrix} v & \theta \end{bmatrix}^T \in H$

$$A_{10}\begin{bmatrix} v \\ \theta \end{bmatrix} = \begin{bmatrix} -\mathbb{P}\big((w_{ss}\cdot\nabla)v + (v\cdot\nabla)w_{ss}\big) \\ -w_{ss}\cdot\nabla\theta - v\cdot\nabla T_{ss} \end{bmatrix} \in X,$$
$$A_{10}^*\begin{bmatrix} v \\ \theta \end{bmatrix} = \begin{bmatrix} \mathbb{P}\big((w_{ss}\cdot\nabla)v - (\nabla w_{ss})^T v - \theta\nabla T_{ss}\big) \\ (w_{ss}\cdot\nabla)\theta + \frac{Gr}{Re^2}\hat{e}_2\cdot v \end{bmatrix} \in X,$$

i.e. $A_{10}, A_{10}^* \in \mathcal{L}(H, X)$. Using (7) and the properties of the trace operator, we see that $B_\Gamma^* \in \mathcal{L}(D((\lambda I - A^*)^{1/2}), \mathbb{R}^2)$. Since B_Ω is bounded, also $B^* \in \mathcal{L}(D((\lambda I - A^*)^{1/2}), U)$. Theory of admissible control and observation operators, see [13, Ch. 4-5], now states that B is an admissible control operator for $\mathbb{T}_A$, thus (6) is a regular linear system. □

This result means that we can use the controllers introduced in [8] to achieve output convergence (5).

3 The Controller

The "main" controller design used in this paper is the low-gain error-feedback controller presented next, and it requires for the semigroup $\mathbb{T}_A$ to be exponentially stable. Consider the error-feedback controller defined by

$$\dot{z}(t) = \mathcal{G}_1 z(t) + \mathcal{G}_2 e(t), \quad z(0) = z_0 \in \mathbb{R}^{2q+1},$$
$$u(t) = Kz(t)$$

with

$$\mathcal{G}_1 = \operatorname{diag}\big(0, G_1^1, \ldots, G_1^q\big), \qquad G_1^k = \begin{bmatrix} 0 & \omega_k \\ -\omega_k & 0 \end{bmatrix},$$
$$\mathcal{G}_2 = \big[-1, G_2^1 \ldots G_2^q\big]^T, \qquad G_2^k = \begin{bmatrix} -1 & 0 \end{bmatrix}^T,$$
$$K = \varepsilon\big[P(0)^\dagger, \mathrm{Re}P(i\omega_1)^\dagger, \mathrm{Im}P(i\omega_1)^\dagger, \ldots \mathrm{Im}P(i\omega_q)^\dagger\big],$$

where $e(t) = y(t) - y_r(t)$ is the regulation error, $P(s) = C(sI - A)^{-1}B$ is the transfer function of (6) and $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse. As shown in [4, 8, 12], there exists $\varepsilon^* > 0$ such that for every $0 < \varepsilon \le \varepsilon^*$ the above low-gain controller achieves the output convergence (5). Note that the controller only requires information of the signal frequencies ω_k and the transfer function values corresponding to the frequencies, and the controller is robust with respect to "small" perturbations on the operators A, B and C.

Depending on the physical parameters and the geometry of Ω, $\mathbb{T}_A$ is not necessarily exponentially stable, c.f [2]. For the unstable case, we could use the more complex observer-based controller introduced in [9, Sec. VI]. This type of controller can also be used for exponentially stable systems to improve the convergence rate in (5), but it comes with the limitation of being infinite-dimensional. However, if we use only the in-domain control inputs u_v and u_θ, the infinite-dimensional controller can be replaced with a system approximation -based finite-dimensional controller introduced in [9]. Furthermore, applicability of the observer-based finite-dimensional controller for boundary control inputs has been studied in [5, 10].

4 A Numerical Example

Let Ω be the unit square and

$$\Gamma_I = \left\{\xi_1 = 0, \frac{5}{8} \le \xi_2 \le \frac{7}{8}\right\}, \quad \Gamma_O = \left\{\xi_1 = 0, \frac{1}{8} \le \xi_2 \le \frac{4}{8}\right\}.$$

We use only two of the presented four control inputs, one on the ξ_1-velocity component within the inlet and

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

one on the temperature within Ω, with the control shape functions

$$b_{v\Gamma}(\xi) = \exp\Big(\frac{-0.00004}{((\frac{5}{8}-\xi_2)(\frac{7}{8}-\xi_2))^2}\Big),$$
$$b_\theta(\xi) = \chi_{[\frac{1}{8},\frac{2}{8}]\times[\frac{1}{16},\frac{3}{16}]}(\xi),$$

where χ denotes the characteristic function, thus $U = \mathbb{R}^2$. Note that even a single control input suffices for output regulation of single output systems as long as $P(\pm i\omega_k)$ is surjective for each frequency ω_k of the reference signal. The observation (4) is characterized by

$$\Omega_y = [\frac{4}{8},\frac{5}{8}]\times[\frac{3}{8},\frac{4}{8}], \quad c_y = \chi_{\Omega_y},$$

and the reference output to be tracked is

$$y_r = 1 - .5\cos(2t).$$

We use the Taylor-Hood finite element spatial dicretization for the velocity part and quadratic (P2) elements for the temperature part of (2) with a uniform triangular grid and triangle edge length 1/16. The steady state solution $x_{ss} = [w_{ss}, T_{ss}]^T$ corresponding to the body force and the heat source

$$f_{bw}(\xi) = \begin{bmatrix} \sin(2\pi\xi_1)\cos(2\pi\xi_2) \\ -\cos(2\pi\xi_1)\sin(2\pi\xi_2)\end{bmatrix},$$
$$f_{bT}(\xi) = \sin(2\pi\xi_1)\cos(2\pi\xi_2)$$

is calculated using the Newton's method.

For the simulation, we use a penalty method, see e.g. [3, Ch. 5.2], to approximately enforce the incompressibility condition $\nabla\cdot v = 0$. The transfer function values at 0 and $\pm 2i$ are calculated using the approximated system operators, and we choose the control tuning parameter $\varepsilon = 0.095$ to roughly maximize the stability margin of the system. For the initial state $[x_0, z_0]^T = [0.5x_{ss}, 0]^T$, the controller achieves accurate output tracking within 80 time units, as depicted in Figure 2.

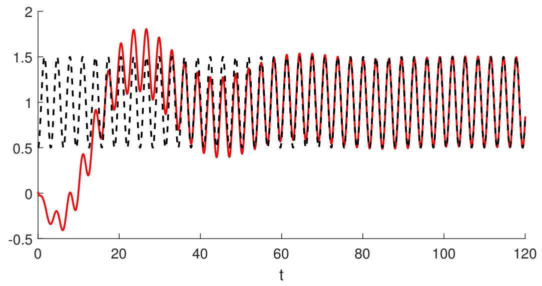


Fig. 2. The system output y (red) and the reference output y_r (black) for $t \in [0, 120]$

5 Conclusion

We studied a temperature output regulation problem for a two-dimensional room model, where the fluid dynamics were modeled by the Boussinesq equations. We considered the Boussinesq equations linearized around a steady state solution and showed that the corresponding abstract linear system is regular. We implemented a simple low-gain controller, which has been considered in several earlier works, to solve the temperature output regulation problem.

Finally, we presented a simulation example for temperature regulation of the room model using the low-gain controller. The controller achieved accurate output tracking thus agreeing with the theoretical results.

Acknowledgement

The research was supported by the Academy of Finland grant number 310489 held by L. Paunonen.

References

[1] Eugenio Aulisa and David Gilliam. *A practical guide to geometric regulation for distributed parameter systems*. Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL, 2016.

[2] John A. Burns, Xiaoming He, and Weiwei Hu. Feedback stabilization of a thermal fluid system with mixed boundary control. *Comput. Math. Appl.*, 71(11):2170–2191, 2016.

[3] M.D. Gunzburger. *Finite Element Methods for Viscous Incompressible Flows: A Guide to Theory, Practice, and Algorithms*. Academic Press, 1989.

[4] Timo Hämäläinen and Seppo Pohjolainen. A finite-dimensional robust controller for systems in the CD-algebra. *IEEE Trans. Automat. Control*, 45(3):421–431, 2000.

[5] Konsta Huhtala, Lassi Paunonen, and Weiwei Hu. Robust output regulation of the linearized Boussinesq equations with boundary control and observation. *arXiv:2010.00960 [math.OC]*., 2020.

[6] Konsta Huhtala, Lassi Paunonen, and Weiwei Hu. Robust output tracking for a room temperature model with distributed control and observation. *arXiv:2007.05313 [math.OC]*., 2020. Accepted for publication in the Proceedings of the 24th International Symposium on Mathematical Theory of Networks and Systems.

[7] Phuong Anh Nguyen and Jean-Pierre Raymond. Boundary stabilization of the Navier-Stokes equations in the case of mixed boundary conditions. *SIAM J. Control Optim.*, 53(5):3006–3039, 2015.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

[8] Lassi Paunonen. Controller design for robust output regulation of regular linear systems. *IEEE Trans. Automat. Control*, 61(10):2974–2986, 2016.

[9] Lassi Paunonen and Duy Phan. Reduced order controller design for robust output regulation. *IEEE Trans. Automat. Control*, 65(6):2480–2493, 2020.

[10] Duy Phan and Lassi Paunonen. Finite-dimensional controllers for robust regulation of boundary control systems. *Math. Control Relat. F.*, 11:95, 2021.

[11] Mythily Ramaswamy, Jean-Pierre Raymond, and Arnab Roy. Boundary feedback stabilization of the Boussinesq system with mixed boundary conditions. *J. Differential Equations*, 266(7):4268–4304, 2019.

[12] Richard Rebarber and George Weiss. Internal model based tracking and disturbance rejection for stable well-posed systems. *Automatica*, 39(9):1555–1569, 2003.

[13] Marius Tucsnak and George Weiss. *Observation and control for operator semigroups*. Birkhäuser Advanced Texts: Basel Textbooks. Birkhäuser Verlag, Basel, 2009.

[14] George Weiss. The representation of regular linear systems on Hilbert spaces. In *Control and estimation of distributed parameter systems (Vorau, 1988)*, volume 91 of *Internat. Ser. Numer. Math.*, pages 401–416. Birkhäuser, Basel, 1989.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Lassi Paunonen*

Rejection of Unknown Harmonic Disturbances in a Boundary Controlled Heat Equation

Abstract: In this short paper we study setpoint control and rejection of unknown harmonic disturbances for a one-dimensional boundary controlled heat equation. The frequencies, amplitudes and phases of the disturbance signal are assumed to be unknown. The designed dynamic error feedback controller consists of an internal model based structure, an adaptive frequency estimator, and an online tuning algorithm for choosing controller parameters.

Keywords: Disturbance rejection, output tracking, heat equation, boundary control, controller design.

***Corresponding Author: Lassi Paunonen:** Mathematics and Statistics, Tampere University, PO. Box 692, 33101 Tampere, Finland, E-mail: lassi.paunonen@tuni.fi. The research is funded by the Academy of Finland grant 310489.

1 Introduction

Partial differential equations are used in modeling of different types of dynamical behaviours such as diffusion processes, waves and vibrations, and fluid flows. These phenomena are also encountered in many areas of engineering and as parts of industrial processes, and this motivates the study of control of models described by partial differential equations. In particular, partial differential equations and their control theory have been used for modeling of flexible links in robotic systems [9], vibration control in oilwell drilling [13], modeling of gas flow in pipe networks [12], and control of traffic flows [3].

In this short paper we study setpoint control and rejection of unknown harmonic disturbances for a one-dimensional boundary controlled heat equation. This partial differential equation model can be used to describe, for example, the time evolution of the temperature profile in a uniform metal rod. The control system defined for $x \in (0,1)$ and $t > 0$ has the form

$$\frac{\partial v}{\partial t}(x,t) = \frac{\partial}{\partial x}\left(c(x)\frac{\partial v}{\partial x}(x,t)\right) + B_d(x)w_{dist}(t) \qquad (1a)$$

$$\frac{\partial v}{\partial x}(0,t) = u(t), \quad \frac{\partial v}{\partial x}(1,t) = 0, \quad v(x,0) = v_0(x) \qquad (1b)$$

$$y(t) = v(x_1,t). \qquad (1c)$$

Here $v(x,t)$ describes the temperature profile at the point $x \in (0,1)$ and at the time $t > 0$ and $c(\cdot)$ is the heat conductivity of the material. Moreover, $u(t)$ is the boundary control input, $y(t)$ is the measured output (the temperature at a single point $x_1 \in [0,1]$), $w_{dist}(t)$ is an external disturbance signal, and $B_d(\cdot)$ is an unknown disturbance input profile.

In this paper we construct a dynamic error feedback controller (Figure 1) such that the output $y(t)$ of the system (1) converges asympotically to a given reference level $y_{ref} \in \mathbb{R}$, i.e.,

$$|y(t) - y_{ref}| \to 0, \qquad \text{as} \quad t \to \infty$$

despite disturbance signals $w_{dist}(t)$ of the form

$$w_{dist}(t) = a_0 + \sum_{k=1}^{q} a_k \cos(\omega_k t + \varphi_k)$$

with unknown amplitudes and phases $\{a_k\}_k, \{\varphi_k\}_k \subset \mathbb{R}$.

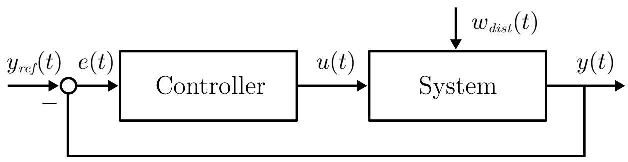


Fig. 1. The dynamic error feedback controller.

If the frequencies $\{\omega_k\}_{k=1}^q$ of $w_{dist}(t)$ are known, then the control problem can be solved using a robust internal model based controller [7, 8, 11]. In this paper we study a situation where the frequencies $\{\omega_k\}_{k=1}^q$ are unknown. Our controller is based on recently introduced general controller design method [1], which consists of an internal model based controller structure, an adaptive estimator for updating the frequencies in the internal model, and an online tuning algorithm for choosing the controller parameters.

For finite-dimensional systems the output tracking and disturbance rejection of harmonic signals with unknown frequencies has been studied earlier in [5, 10, 14, 18]. For distributed parameter systems and partial differential equations this problem has only been studied in very few references either for particular models [2] or under fairly restrictive assumptions on the controlled

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

system [16, 17]. Some of the main benefits of the controller structure introduced in [1] are that the frequency estimation method can be chosen quite freely and output tracking and disturbance rejection are achieved under fairly minimal assumptions on the system and the frequency estimation method.

2 Controller Design

The problem of output tracking in the presence of an external disturbance signal with unknown frequencies can be solved with the controller introduced in [1]. The controller consists of the following parts:

- A general internal-model based controller structure
- An adaptive estimator for forming (time-dependent) estimates $\{\hat{\omega}_k(t)\}_k$ of the unknown frequencies $\{\omega_k\}_k$ of the disturbance $w_{dist}(t)$
- An online tuning algorithm for updating the frequencies of the internal model and the gain parameters of the controller at regular intervals.

The main prequisites for the controller design are that the plant is stabilizable and detectable (in a suitable sense) and that it does not have transmission zeros in the neighbourhoods of the points $\{\pm i\omega_k\}_k \subset \mathbb{C}$. Both of these conditions are satisfied for the heat equation (1).

The general controller structure introduced in [1] can be used together with any adaptive frequency estimator which is capable of identifying unknown frequencies from a continuous-time signal. One suitable multi-frequency estimator is introduced in the reference [4], and we use this estimator in the simulations in Section 3. Other possible choices of adaptive frequency estimators have been introduced in [6] and [15]. In the estimation we assume the number of frequencies in the disturbance signal to be known (some estimation algorithms offer online identification of the number of frequencies).

For the heat equation model (1) the final controller is a nonautonomous coupled system consisting of a linear partial differential equation part (an observer-type part for the system (1)), a linear ordinary differential equation part (the time-dependent internal model), and a nonlinear ordinary differential equation part (the nonlinear adaptive frequency estimator). The tuning algorithm updates the frequencies of the internal model and the gain parameters of the controller at regular update intervals. The frequencies are not updated if at the given update time the frequency estimates provided by the adaptive estimator are not valid (frequency estimates are not positive real numbers), or if the frequency estimates are too close to each other (in which case closed-loop stability is difficult to achieve). The details of the controller structure and the controller tuning algorithm are presented in [1]. The results in [1] guarantee the convergence of the output $y(t)$ to the reference level y_{ref} despite the external disturbance signal $w_{dist}(t)$ provided that the frequency estimates formed by the adaptive estimator converge to the true frequencies of the disturbance signal. This convergence property is theoretically guaranteed for the estimator in [4], but in practice this property depends on the choice of the estimator's parameters (see the discussion in Section 4).

3 Simulation Results

In this section we illustrate the controller performance with numerical simulations. In the simulation we assume constant heat conductivity $c(x) \equiv 1$ for $x \in [0,1]$ and let the output $y(t)$ be the temperature measurement at the endpoint $x_1 = 1$. The gain parameters γ_1 and γ_2 of the frequency estimator in [4] are chosen as $\gamma_1 = 0.005$ and $\gamma_2 = 10$. The disturbance input profile is chosen as $B_d(x) = \sin(3.5x)$ for $x \in [0,1]$ (this information is not used in the controller design). The disturbance input is

$$w_{dist}(t) = \cos(1.5t - 0.4) + 5\cos(5t + 0.2)$$

and thus has two frequencies $\omega_1 = 1.5$ and $\omega_2 = 5$. The reference level of the system is chosen to be $y_{ref} = 56$. The controller frequency update times are $t = 0, 1, 2, \ldots$.

In the simulations the heat equation (1) is approximated with the Finite Difference scheme with $N = 100$ points. Figures 2 and 3 depict the controlled output and the evolution of the controlled temperature profile on the interval $0 \le t \le 15$ with the initial state $w_0(x) \equiv 50$ of the system (1). In addition, Figure 4 depicts the convergence of the frequency estimates.

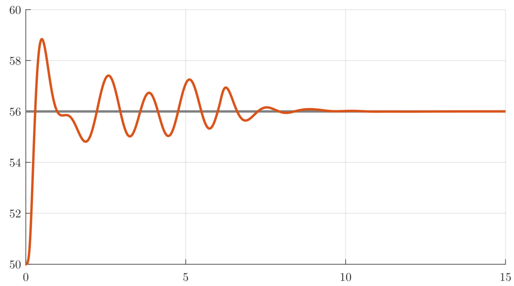


Fig. 2. The output of the controlled heat equation.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

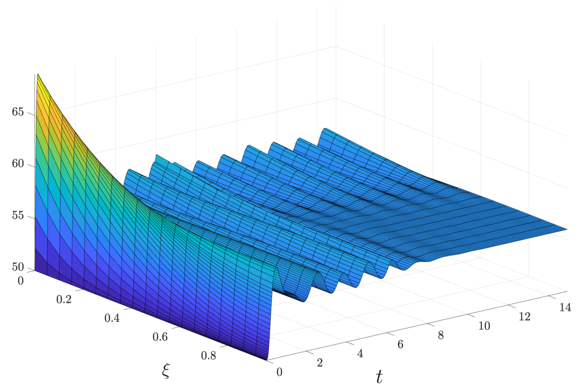


Fig. 3. The temperature profile of the controlled heat equation.

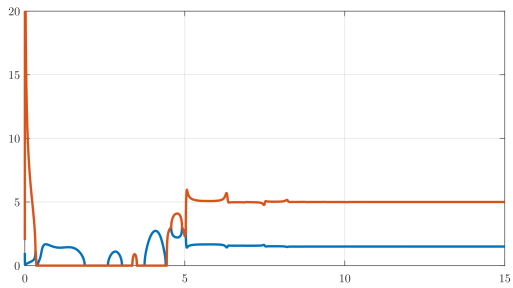


Fig. 4. The convergence of the frequency estimates.

In the case where the frequencies $\{\omega_k\}_k$ of the disturbance are known, the internal model based controllers are robust in the sense that they achieve the setpoint tracking and disturbance rejection even in the presence of small uncertainties or perturbations in some of the parameters of the control system (1) [11]. Also in the case of unknown frequencies, the controller design in [1] is robust in the sense that if the parameters of the system (1) contain small uncertainties, then the setpoint tracking is achieved *approximately* with a small residual tracking error. However, robustness margin can become very small due to the fact that especially the convergence of the frequency estimates $\{\hat{\omega}_k(t)\}_k$ to the correct frequencies $\{\omega_k\}_k$ can be very sensitive to perturbations in the parameters of the system (1).

We illustrate these robustness properties by considering a perturbed control system with an additional unmodeled reaction term. More precisely, we use our controller designed for the system (1) (with $c(x) \equiv 1$ and $x_1 = 1$) to control the system

$$
\begin{aligned}
&\frac{\partial v}{\partial t}(x,t) = \frac{\partial^2 v}{\partial x^2}(x,t) + g(x)v(x,t) + B_d(x)w_{dist}(t)\\
&\frac{\partial v}{\partial x}(0,t) = u(t), \quad \frac{\partial v}{\partial x}(1,t) = 0, \quad v(x,0) = v_0(x)\\
&\quad y(t) = v(1,t).
\end{aligned}
$$

The additional reaction term $g(x)v(x,t)$ can be considered as a perturbation to the original system (1) when the values of the profile function $g : [0,1] \to \mathbb{R}$ are small.

In the simulations we let $g(x) = \beta(\cos(\pi x) + 0.25)$ with $\beta \in \mathbb{R}$. The other simulation parameters have the same values as in the beginning of the section. If $|\beta|$ is sufficiently small, the controller should (theoretically) achieve convergence of the output $y(t)$ to the reference level $y_{ref} = 56$ approximately in the sense that for some small $\varepsilon > 0$ and for some $t_0 > 0$ the output satisfies $|y(t) - y_{ref}| \leq \varepsilon$ for all $t \geq t_0$.

Figure 5 depicts the controlled output and the frequency estimates with the perturbation parameter $\beta = 2 \cdot 10^{-5}$. The results illustrate that even with such a very small perturbation, the output tracking has a residual error, and the the frequency estimates no longer converge to the correct frequencies $\omega_1 = 1.5$ and $\omega_2 = 5$ of $w_{dist}(t)$. The convergence of the frequency estimates would guarantee that $y(t)$ converges to the reference level y_{ref}. However, in the case of perturbations in the system (1), the signal used in the frequency estimation contains non-decaying noise which makes the estimation of the frequencies more difficult.

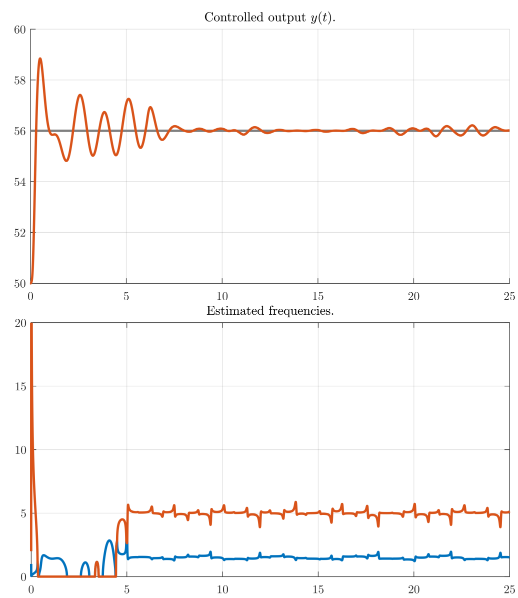


Fig. 5. The output and the frequency estimates with $\beta = 2 \cdot 10^{-5}$.

For larger values of β the frequency estimation can fail completely and consequently the output will fail to converge to the reference level y_{ref} even approximately. This is illustrated by Figure 6 which depicts output and

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

the frequency estimates in the case of the perturbation parameter $\beta = 0.001$.

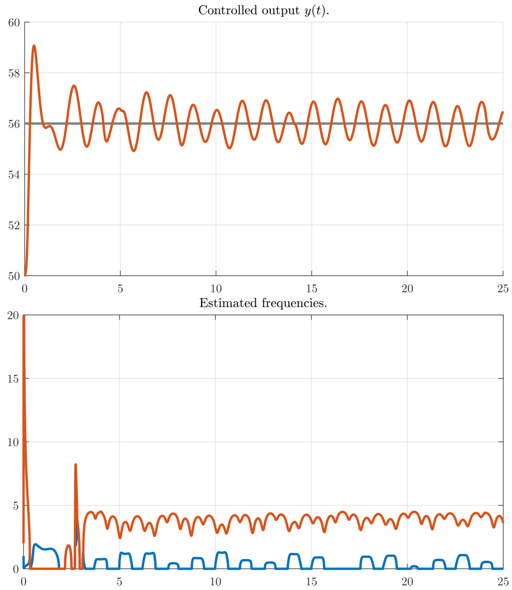


Fig. 6. The output and the frequency estimates with $\beta = 0.001$.

We note that the convergence of the output and the frequency estimates also depend heavily on the shape of the profile function $g : [0,1] \to \mathbb{R}$, and not only its size. In particular, output tracking and frequency estimation can still be achieved successfully for selected profile functions with much larger values, and on the other hand the output tracking and frequency estimation can fail for certain profile functions with even smaller values than in the above simulation examples.

The gain parameters γ_1 and γ_2 of the frequency estimator were constant in the simulations and their effect on the frequency estimation under perturbations of the system (1) was not investigated. Modifying these parameters could potentially help in reducing the oscillatory behaviour of the frequency estimation under perturbations of the system.

4 Conclusions

In this short paper we have studied controller design for setpoint tracking of a boundary controlled heat equation in the presence of an external disturbance with unknown frequencies. As illustrated by the simulation results, the controller achieves satisfactory performance in the control of the nominal system when the disturbance signal has two unknown frequencies.

In the case where the control system contains uncertainty in the form of a small unmodeled reaction term, our simulation results illustrate that the controller will no longer achieve perfect asymptotic tracking, and the tracking can fail completely as a result of even very small perturbations. The success of the output tracking and the disturbance rejection depend heavily on the success of the frequency estimation. In practice, the uncertainties and perturbations in the parameters of the system cause model mismatch between the observer-part of the controller and the actual control system, and as a result the frequency estimation is carried out based on a signal which contains the unknown frequency components, an exponentially decaying transient part, and a persistent error. This persistent error can very easily cause the adaptive frequency estimator to fail in identifying the correct frequencies. Even with estimators which are *input-to-state stable* (the estimator [4] used in the simulations has this property) even small model mismatch can be cause large errors in the frequency estimation, and as a result the robustness margin of the full controller will be very limited.

Because of the strong influence of the frequency estimation on the achievability of the setpoint tracking, the choice of the frequency estimator can have a big impact on the final robustness margin of the controller. In particular, the frequency estimator's capability of handling persistent and unstructured noise in the estimation signal will be beneficial to the performance of the full controller under perturbations of the control system. The frequency estimator in [4] which was used in the simulations was designed for estimation of frequencies from a signal which is a pure combination of distinct frequency components, and the fact that the estimation is achieved approximately even under persistent error signals is a commendable additional property of this estimator.

References

[1] S. Afshar and L. Paunonen. Time-varying internal models in robust output regulation. *arXiv e-prints*, February 2020. Available at https://arxiv.org/abs/2002.09684

[2] H. Anfinsen, T. Strecker, and O. M. Aamo. Rejecting unknown harmonic disturbances in 2×2 linear hyperbolic PDEs. *IEEE Trans. Control Syst. Technol.*, 25(6):1935–1946, 2017.

[3] M. Burkhardt and Huan Yu, and M. Krstic. Stop-and-go suppression in two-class congested traffic. *Automatica J.*

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

IFAC, 125: 109381, 8, 2021.

[4] D. Carnevale and A. Astolfi. A minimal dimension observer for global frequency estimation. In *2008 American Control Conference*, pages 5236–5241, Seattle, WA, USA, Jun 11–13 2008.

[5] D. Carnevale, S. Galeani, M. Sassano, and A. Astolfi. Robust hybrid estimation and rejection of multi-frequency signals. *Internat. J. Adapt. Control Signal Process.*, 30(12):1649–1673, 2016.

[6] B. Chen and G. Pin and Wai M. Ng, Shu Yuen Hui, and T. Parisini. An adaptive-observer-based robust estimator of multi-sinusoidal signals *IEEE Trans. Automat. Control*, 63(6):1528–1541, 2018.

[7] T. Hämäläinen and S. Pohjolainen. Robust regulation of distributed parameter systems with infinite-dimensional exosystems. *SIAM J. Control Optim.*, 48(8):4846–4873, 2010.

[8] Jie Huang. *Nonlinear Output Regulation, Theory and Applications*. SIAM, Philadelphia, 2004.

[9] A. Macchelli and C. Melchiorri and S. Stramigioli. Port-based modeling of a flexible link. *IEEE Trans. Robot.*, 23(4):650–660, 2007.

[10] R. Marino and P. Tomei. Output regulation for linear systems via adaptive internal model. *IEEE Trans. Automat. Control*, 48(12):2199–2202, 2003.

[11] L. Paunonen. Controller design for robust output regulation of regular linear systems. *IEEE Trans. Automat. Control*, 61(10):2974–2986, 2016.

[12] F. Rüffler, V. Mehrmann, and F. M. Hante. Optimal model switching for gas flow in pipe networks *Netw. Heterog. Media*, 13(4): 641–661, 2018.

[13] B. Saldivar and S. Mondié and S.-I. Niculescu and H. Mounier and I. Boussaada. A control oriented guided tour in oilwell drilling vibration modeling. *Annu. Rev. Control*, 42:100–113, 2016.

[14] A. Serrani, A. Isidori, and L. Marconi. Semiglobal nonlinear output regulation with adaptive internal model. *IEEE Trans. Automat. Control*, 46(8):1178–1194, 2001.

[15] A. Vediakova and A. Vedyakov, A. Pyrkin, A. Bobtsov, and V. Gromov. Frequency Estimation of Multi-Sinusoidal Signals in Finite-Time. *arXiv e-prints*, September 2020. Available at https://arxiv.org/abs/2009.06400

[16] Xinghu Wang, Haibo Ji, and Jie Sheng. Output regulation problem for a class of SISO infinite dimensional systems via a finite dimensional dynamic control. *J. Syst. Sci. Complex.*, 27(6):1172–1191, 2014.

[17] Xinghu Wang, Haibo Ji, and Chuanrui Wang. Output regulation for a class of infinite dimensional systems with an unknown exosystem. *Asian J. Control*, 16(5):1548–1552, 2014.

[18] Dabo Xu and Jie Huang. Robust adaptive control of a class of nonlinear systems and its applications. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 57(3):691–702, 2010.

Ari Kuisma

Automaatiojärjestelmän modernisointi ja etätestaus

Tiivistelmä: Jyväskylän ammattikorkeakoulussa oli tarve modernisoida kaksi opetuskäytössä olevaa Metso DNA -automaatiojärjestelmää. Toinen järjestelmistä alkoi olla jo elinkaarensa lopussa ja toinenkin vaati päivittämistä.

Modernisoinnin lähtökohdaksi otettiin järjestelmän etäkäytettävyys. Opiskelijoiden työskentely ei saisi enää olla rajoitettu koulun tiloihin, ja järjestelmän pitäisi toimia ilman erillisten sovellusten asentamista omalle tietokoneelle. Samalla haluttiin yhdistää kaksi erillistä automaatiojärjestelmää yhdeksi järjestelmäksi. Koska kyseessä oli modernisointiprojekti, neuvotteluja käytiin ainoastaan Valmetin kanssa uuden Valmet DNA:n ominaisuuksista.

Valmet tarjosi ratkaisua, jossa kaikki vaatimukset saatiin toteutettua. Koulun palvelintiloihin asennettiin yksi uusi palvelin, Private Cloud, johon aiemmin erillisiä työasemia ja palvelimia vaatineet koneet asennettiin virtuaalikoneina. Tähän Private Cloudiin otetaan yhteys Citrixin etätyökaluilla niin koulun sisältä kuin sisäverkon ulkopuoleltakin.

Lopputuloksena saatiin toimiva järjestelmä, jolla sovellusohjelmien suunnittelu onnistuu omilta koneilta, myös koulun ulkopuolelta. Järjestelmää määriteltäessä ei osattu vielä aavistaa, miten suuri tarve uusille ominaisuuksille tulee covid-19-pandemian vuoksi olemaan.

Lukukaudella 21/22 lähes kaikki opiskelijat ovat tehneet sekä suunnittelua että testausta koulun ulkopuolelta. Etätestaus on hoidettu niin, että vähintään yksi henkilö on ollut laboratoriossa valvomassa prosessin toimintaa. Testaaja on voinut olla etänä ja yhteys laboratorioon on otettu videopuhelun avulla. Videokuvaa on käytetty myös apuna havainnoimaan prosessin tilaa. Opiskelijat ovat ladanneet oman sovellusohjelman järjestelmään ja ovat tehneet sekä I/O-testin että toiminnallisen testauksen etänä.

Avainsanat: automaatiojärjestelmä, virtuaalikone, Private Cloud, etäkäyttö, Valmet DNA

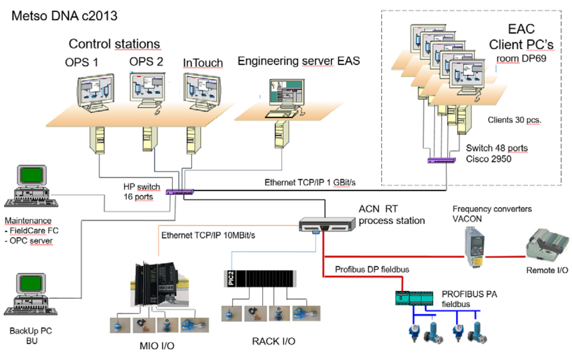
Ari Kuisma: Jyväskylän ammattikorkeakoulu, automaatiotekniikan lehtori, S-posti: ari.kuisma@jamk.fi

1 Johdanto

Opiskelun monimuotoistuminen on luonut kasvavan tarpeen opetuksessa käytettävien järjestelmien etäkäytettävyydelle. Aikaisemmin on ollut aivan normaalia, että opetus on tapahtunut ainoastaan tietyssä tilassa ja tiettyyn aikaan. Yksittäisen opetuskerran vuoksi on matkustettu pitkiäkin matkoja. Nykyinen tekniikka mahdollistaa kuitenkin yhä paremmin myös laitteiden etäkäytön, jota haluttiin tuoda Jyväskylän ammattikorkeakoulussa (JAMK) myös automaatiojärjestelmien opetukseen. Tavoitteena automaatiojärjestelmän modernisoinnissa oli se, että järjestelmää voisi käyttää täysin identtisesti niin koulun sisäverkosta kuin koulun ulkopuoleltakin.

2 Järjestelmän rakenne

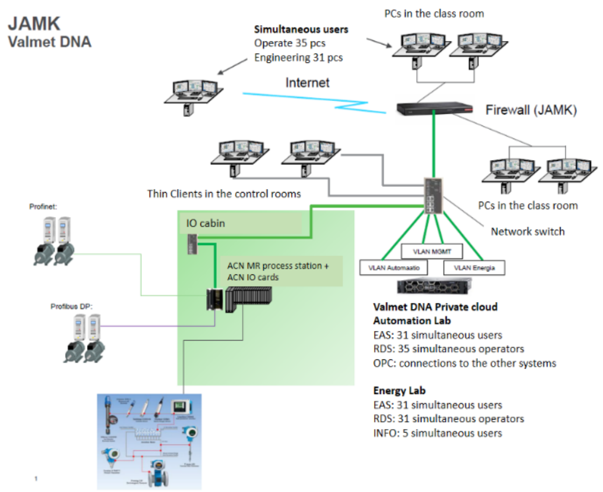
Aikaisempi Metso DNA -järjestelmä oli perinteinen hajautettu järjestelmä, jossa jokaista työasemaa tai palvelinta varten oli varattu oma tietokone. Kuvassa 1 nähdään automaatiolaboratoriossa olleen järjestelmän rakenne.



Kuva 1. Automaatiolaboratorion vanha Metso DNA -järjestelmä

Järjestelmään oli kytketty yhteensä 38 tietokonetta ja palvelinta, joissa kaikissa piti olla asennettuna tietyt Metso DNA:n ohjelmistot. Tällainen järjestelmärakenne aiheutti paljon työtä ylläpidolle, etenkin kun samoja työasemia käytettiin myös monen muun suunnittelujärjestelmän opetuksessa. Oli varsin tyypillistä, että työasemiin tuli automaattinen päivitys, joka esti tai häiritsi Metso DNA:n ohjelmistojen toimintaa.

Uuden Valmet DNA -järjestelmän suuntaa antava periaatekuva on esitetty kuvassa 2. Tästä kuvasta puuttuvat energialaboratorion laitteistot. JAMKin palvelinhuoneeseen asennettiin yksi palvelin, joka toimii järjestelmän ”yksityisenä pilvenä”, Private Cloudina. Kaikki aikaisempien järjestelmien vastaavat työasemat ja palvelimet ovat nyt asennettuna palvelimeen virtuaalikoneina, ja Private Cloud on ainoa tietokone, jossa on Valmet DNA -ohjelmistoja asennettuna. Enää ei siis tarvita yhtä erillistä luokkahuonetta ohjelmistoineen sovellussuunnittelua varten, vaan suunnittelua voidaan tehdä mistä vaan. Kaikki yhteydet järjestelmään hoidetaan verkko-selaimella Citrix-etätyöpöydän kautta.



Kuva 2. Uusi Valmet DNA -järjestelmä

Aikaisemmassa toteutuksessa JAMKissa oli kaksi täysin erillistä Metso DNA -järjestelmää, joilla ei ollut mitään yhteyttä keskenään. Nyt molempien laboratorioiden prosessiohjaimet (PCS) on kytketty samaan Private Cloudiin yhteisellä virtuaalilähiverkolla (VLAN). Järjestelmää varten ei vedetty uutta kaapelointia, vaan Valmet DNA -järjestelmä toimii JAMKin lähiverkon sisällä omassa VLAN:ssa.

Automaatiolaboratoriossa uusittiin myös prosessiohjain. Vanhan, turhankin tehokkaan, ACN RT:n tilalle vaihdettiin uusi aiempaa pienempi ACN MR -prosessiohjain. Energialaboratoriossa säilytettiin entiset kaksi ACN CS -prosessiohjainta. Kummassakaan laboratoriossa prosessiohjaimista kentän suuntaan ei tehty muutoksia.

Tällä hetkellä ohjattavina prosesseina energialaboratoriossa ovat neljästä erillisestä moduulista koostuva vesi- ja höyryprosessi, jossa moduuleita voidaan ajaa joko yksikköprosesseina tai yhtenä kokonaisuutena. Automaatiolaboratoriossa ohjattavana prosessina on kolmen säiliön vesiprosessi suljetulla vesikierrolla. Kaikkien prosessien kenttälaitteet on kytketty prosessiohjaimiin Profibus DP -kenttäväylillä.

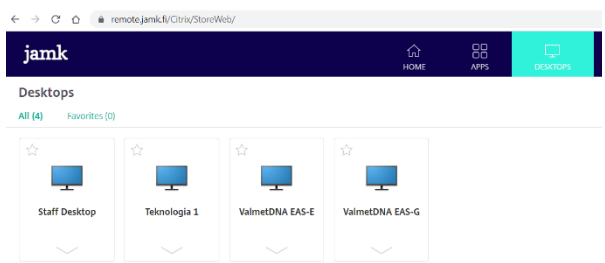
3 Käyttäjien hallinta

Kirjautuminen Valmet DNA -järjestelmään tapahtuu täysin identtisesti sekä JAMKin sisäverkosta että koulun ulkopuolelta. Tällä järjestelyllä saatiin käyttäjä-kokemuksesta mahdollisimman yksinkertainen. Toisin kuin vanhassa järjestelmässä, nyt käyttäjä tarvitsee koneelle ainoastaan verkkoselaimen ja internetyhteyden.

Yhtä lailla järjestelmään kirjaudutaan aina samalla tavalla, on sitten tarkoitus tehdä sovellussuunnittelua tai operoida prosessia. Yhteys Private Cloudiin päätettiin toteuttaa Citrixin etätyöpöydällä, joka oli jo valmiiksi JAMKilla käytössä.

Automaatiojärjestelmä jaettiin kahteen toiminnalliseen osaan: E- ja G-järjestelmään. E-järjestelmä on energiatekniikan laboratoriota ja G-järjestelmä automaatiotekniikan laboratoriota varten. Kaikki toiminnot voisivat olla myös yhden virtuaalikoneen takana, mutta tällä järjestelyllä saatiin pienennettyä riskiä odottamattomista käynnistyksistä.

Kuvassa 3 on esitetty henkilökuntatunnuksilla avattu etätyöpöytä, josta voidaan edelleen avata joko E- tai G-järjestelmä. Opiskelijat harjoittelevat sovellussuunnit-telua ja testausta pääsääntöisesti G-järjestelmällä. E-järjestelmä on varattu energiaprosessin operointia varten.



Kuva 3. Virtuaalikoneen avaus

Citrixin etätyöpöytä on käytössä jo valmiiksi kaikilla opiskelijoilla ja henkilökunnalla, ja kirjautumiseen riittävät normaalit opiskelija- ja henkilökunta-tunnukset. Kuvassa 3 näkyvät virtuaalikoneet (ValmetDNA EAS-E, ValmetDNA EAS-G) vaativat kuitenkin oman käyttöoikeutensa, joka myönnetään jokaiselle käyttäjälle erikseen. Käyttöoikeus liitetään olemassa olevaan käyttäjätunnukseen.

ISBN: 13 978-952-5183-59-7

Tällä hetkellä käyttäjiä hallitaan siten, että opettaja lähettää listan halutuista käyttäjistä tietohallinnon edustajalle, joka antaa oikeudet Valmet DNA -virtuaalikoneille. Tämä on osoittautunut melko vaivattomaksi, koska opintojakson alussa ilmoittautumisjärjestelmästä saatava raportti mahdollistaa lisäkäyttöoikeuksien myöntämisen virtuaalikoneille massa-ajona. Jälki-ilmoittautuneita yksittäisiä käyttäjiä on tarvinnut lisätä vain harvoin. Koska järjestelmä on ollut vasta vähän aikaa käytössä, käyttäjäoikeuksien poistolle ei ole ollut vielä tarvetta, eikä sille siten ole vielä luotu käytäntöä.

Käyttäjien oikeuksia ei ole rajoitettu ohjelmallisesti, vaan kaikki toiminnan rajoitukset on hoidettu ainoastaan ohjeistuksilla. Opiskelijat opastetaan käyttämään järjestelmää samalla kunnioituksella kuin käyttäisivät ajossa olevan tehtaan automaatiojärjestelmää. Mikään ei estä ohjelmallisesti käynnistämästä prosesseja etänä tai poistamasta järjestelmästä sovellusohjelmia. Prosessien odottamaton käynnistäminen on estetty kytkemällä prosessien turvakytkimet auki silloin, kun ne eivät ole käytössä. Prosesseja saa ajaa etänä vain, jos joku on laboratoriossa valvomassa ajoa.

4 Testaus ja simulointi

Sovellussuunnittelu ei poikkea aikaisemmasta järjestelmästä, joten se jätetään tässä artikkelissa käsittelemättä. Sovellussuunnittelun kannalta tärkein muutos vanhaan järjestelmään on se, että nykyään sovellussuunnittelua voi tehdä ajasta ja paikasta riippumatta.

Testausta tehdään kahdella tavalla, joko fyysisellä prosessilla tai fyysisen prosessin simulaattorilla. Periaatteena on, että opiskelijan on tehtävä aina I/O-testaus oikealla prosessilla. Silloin opitaan vähintään alkeet ongelmanratkaisusta todellisessa ympäristössä, jossa vika voi olla ohjelmiston lisäksi myös sähköisissä kytkennöissä, paineilmassa tai mekaanisella puolella. Ensimmäisissä testauksissa opettaja on ollut pääsääntöisesti valvomassa testausta, mutta myöhäisemmässä vaiheessa opiskelijat ovat testanneet myös keskenään ilman opettajaa. Testauksien aikana vähintään yhden henkilön on oltava aina laboratoriossa seuraamassa testauksen etenemistä. Vaikka ohjattavassa vesiprosessissa on suljettu vesikierto, on kuitenkin mahdollista, että vesi ajetaan säiliön ylivuodosta yli, tai pumppuja ajetaan kuivana, jos lukitukset eivät toimi oikein. Prosessiin ei ole tehty mitään kovan puolen lukituksia, vaan kaikki toiminnallisuus on opiskelijan itse tekemää. I/O-testauksen jälkeen opiskelija voi jatkaa sovelluskehitystä ja tehdä seuraavat testaukset pelkästään simulaattorilla.

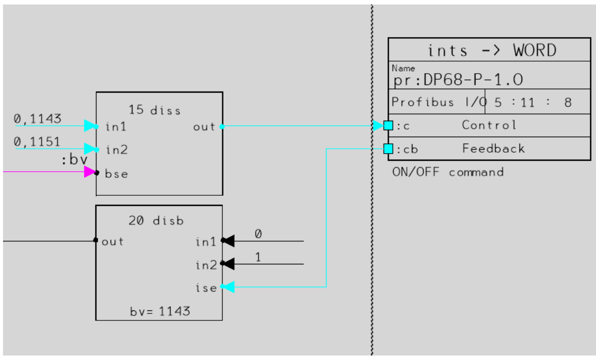
Vesiprosessin simulaattori teetettiin vuonna 2019 opinnäytetyönä. Simulaattorin rakenteeseen on mahdollista tutustua tarkemmin Valkeejärven opinnäytetyön raportissa [1]. Alun perin simulaattori teetettiin helpottamaan vesiprosessin testauksen ruuhkaa. Silloin ei osattu aavistaa, miten suuressa roolissa simulaattori olisi covid-19-pandemian aikana, kun suurin osa opiskelijoista tekee testausta etänä.

Opiskelijan ei tarvitse ymmärtää simulaattorin sisäistä toimintaa. Riittää, että hän muuttaa sovellusohjelman I/O-liitännät viittauksiksi simulaattoriin. Kuvassa 4 on otos automaatiolaboratorion vesiprosessin I/O-listasta, jossa näkyvät ensin Profibus DP:n osoitteet ja niitä vastaavat simulaattorin osoitteet.

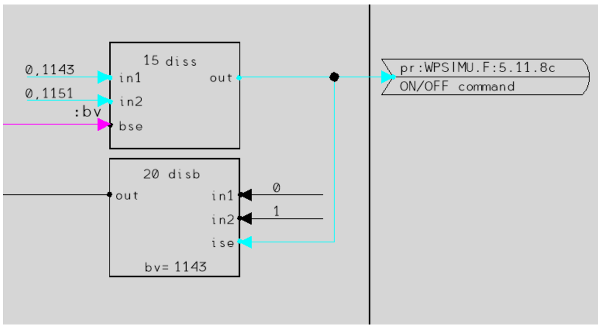
Profibus DP				Simulator	
FBC	Slave	Offset	Scale	Address	Data type
5	7	0	Real + ST	pr:WPSIMU.F:5.7.0m	ana
5	9	0	Real + ST	pr:WPSIMU.F:5.9.0m	ana
5	5	0	Real + ST	pr:WPSIMU.F:5.5.0c	ana
5	8	0	Real + ST	pr:WPSIMU.F:5.8.0c	ana
5	11	8	1143, 1151 WORD	pr:WPSIMU.F:5.11.8c	ints
5	11	8.4	bit	pr:WPSIMU.F:5.11.8.4m	bin
5	11	10	0…10000 WORD	pr:WPSIMU.F:5.11.10c	ints
5	11	16	word	pr:WPSIMU.F:5.11.16m	ints
5	11	14	ints	pr:WPSIMU.F:5.11.14m	ints
5	12	8	1143, 1151 WORD	pr:WPSIMU.F:5.12.8c	ints
5	12	8.4	bit	pr:WPSIMU.F:5.12.8.4m	bin
5	12	10	0…10000 WORD	pr:WPSIMU.F:5.12.10c	ints

Kuva 4. Ote vesiprosessin I/O-listasta

Kuvassa 5 on puolestaan esimerkkinä taajuusmuuttajan ON/OFF-ohjaus Profibus DP:n kautta. Kuvassa 6 esitetään vastaava ohjaus, kun ohjaus käännetään simulaattorille. Kun opiskelija I/O-testauksen jälkeen haluaa jatkaa testausta simulaattorilla, Profibus-korttien paikalle vaihdetaan tulo- sekä lähtöpuolella niitä vastaavat viittaukset simulaattoriin. Sovellusohjelmaan ei tarvitse tehdä muita muutoksia.



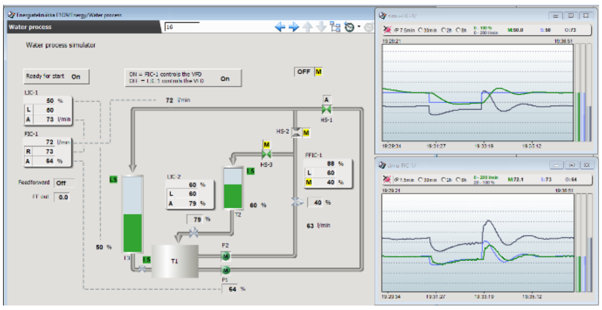
Kuva 5. Profibus-ohjaus taajuusmuuttajalle



Kuva 6. Simulaattoriohjaus taajuusmuuttajalle

Opiskelijoiden käytössä on varauskalenteri, josta he voivat varata itselleen testausaikaa joko fyysisellä vesiprosessilla tai simulaattorilla. Tällä hetkellä opiskelijoiden käytössä on vain yksi simulaattori, mutta jatkossa simulaattoreita on tarkoitus ottaa käyttöön useampia.

Kuvassa 7 on esitetty kuvakaappaus kaskadisäätimen testauksesta simulaattorin avulla. Prosessin dynamiikat on mallinnettu simulaattoriin niin tarkasti, että säätöpiirit voidaan virittää simulaattorin avulla. Sen jälkeen samoilla parametreilla voi ajaa oikeaa fyysistä prosessia.



Kuva 7. Kaskadisäätimen testaus simulaattorilla

Testausta on tehty etäopiskeluaikana siten, että kaikilla testaukseen osallistuvilla on ollut yhteinen videoneuvottelu auki. Testausta tekevä opiskelija on vuorollaan jakanut oman näyttönsä, jolloin muut ovat voineet seurata testauksen etenemistä. Testausta valvova opettaja on voinut seurata ja opastaa testausta videoneuvottelun avulla. Myös vianetsintää on tehty yhdessä opiskelijoiden kanssa videoneuvottelun välityksellä.

5 Käyttökokemuksia

JAMKissa opetetaan automaatiojärjestelmiä Sähkö- ja automaatiotekniikan sekä Energia- ja ympäristötekniikan koulutusohjelmissa. Molemmissa koulutusohjelmissa on tarjolla normaalin päiväopetuksen lisäksi monimuoto-ohjelmat. Sähkö- ja automaatiotekniikan koulutusohjelmassa opetukseen osallistuu joka vuosi myös muutamia vaihto-opiskelijoita.

Uusi Valmet DNA -automaatiojärjestelmä asennettiin elokuussa 2020 ja se on ollut opetuskäytössä syyskuun alusta alkaen. Pelkästään automaatiojärjestelmän tarpeisiin varattu tehokas palvelin on riittänyt erinomaisesti tähän tarkoitukseen. Käyttäjän kannalta pienikin hitaus virtuaalikoneiden toiminnassa olisi turhauttavaa. Puolen vuoden käyttökokemus on osoittanut järjestelmän toimivan viiveettä.

Yhtäaikaisten käyttäjien lisenssien määrä on osoittautunut riittäväksi. Valmet DNA -lisenssejä on käytössä vanhan järjestelmän mukaisesti vain luokkahuoneelliselle opiskelijoita, eli noin neljännekselle käyttäjien määrästä. Ennen järjestelmän käyttöä rajoittivat myös oppilaitoksen aukioloajat, koska se oli käytettävissä vain yhdessä luokassa. Nyt aukioloajat rajoittavat ainoastaan fyysisen prosessin testausta. Virtuaalikoneiden myötä järjestelmän käyttö on mahdollista ympäri vuorokauden jokaisena viikonpäivänä, ja opiskelijat näyttävät hyödyntävän tätä mahdollisuutta tehokkaasti.

Opiskelijoiden palaute uudesta järjestelmästä sekä sen mahdollistamasta joustavasta opiskelusta on ollut pelkästään positiivista. Järjestelmällä on tällä hetkellä yli sata käyttäjää, joista lähes kaikki ovat tehneet sekä sovellussuunnittelua että testausta pelkästään etänä. Myös opettajien näkökulmasta järjestelmä on toiminut juuri kuten sen haluttiinkin toimivan, ja jopa odotuksia paremmin.

Viitteet

[1] Valkeejärvi. M., Vesiprosessin mallintaminen ja simulaattorin rakentaminen, opinnäytetyö AMK, Jyväskylän ammattikorkeakoulu, Jyväskylä, 2019 Julkaisun pysyvä osoite: http://urn.fi/URN:NBN:fi:amk-2019053113790

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

Kai Zenger*

Digitalization and automation in teaching

Abstract

Digitalization is a magic buzzword in almost all disciplines in modern society. It does not only concern technology, but life in general, and without doubt the key idea today is linked to digital solutions found everywhere in the world. The technological revolution has developed in enormous steps and is closely connected to such concepts as game industry, machine learning, artificial intelligence, robotics, the Internet of Things, intelligent manufacturing, quantum computing etc. Even though the youth of today have grown to be "diginative" in many aspects, the teaching and learning of the more and more challenging technological solutions has become, not easier but even more challenging. Digitalization in teaching and learning has many conflicting aspects and consequences, and the current era of covid 19 pandemia is one clear challenge to demonstrate the needs, possibilities and problems in teaching. The paper discusses some of the topics related - especially but not only - concentrating in automation.

Keywords: Education, Digitalization, Automation, Control

***Corresponding Author: Kai Zenger:** Aalto University, Department of Electrical Engineering and Automation, E-mail: kai.zenger@aalto.fi

1 Introduction

The needs for advanced engineering education is enormous globally. For example, the environmental problems set large new specifications and demands for engineering, and the solutions e.g. in resisting the destructive climate change need more complicated technical solutions than before. It is well-known that saving energy and reducing emissions in engine and power systems can happen but only by technically advanced and unfortunately also complicated solutions. One example of that are the combustion engines that drive cars and ships. To reduce nitrogen oxide and green house gas emissions is possible by introducing new combustion methods. However, these are extremely complex systems that need exact modelling and control. In other words: Automatic control plays a key role.

The covid-19 caused a major change in teaching from the early 2020. The *forced digitalization jump* that has been anticipated in teaching before now got a quick start, facing lots of unprepared teachers. Even though web-based teaching portals have been effectively used for quite a long time already, the teaching itself has mostly been based on lectures, exercise hours, seminars, laboratory exercises, homework etc. Changing all this for remote teaching did not turn out to be straightforward at all, and in practice each teacher has been obliged to act in the best way that he/she has been able to. For example, in Aalto University the instructions from the education administration and pedagogic experts have been quite general leaving much room for interpretations, and have therefore not been really helpful in practical teaching. All this is somehow understandable, because the change to distant learning happened so fast and unexpectedly. However, now that time has passed it would be good to have a more strategic and common approach to digital teaching in the university level.

In Automatic Control the work to this end can be initiated by looking of some earlier reports on the challenges in education, and mapping these to the current situation. In [1] the educational challenges of especially control engineering were discussed, the main question being in approaching the learning disciplines of the young generation, especially in mathematics and physics. Of equal but maybe even more important issue today was discussed in [2] where the future and threats towards the curriculum in Automation and Control were anticipated. Briefly, the main questions were ([2])

– Is automation losing its position as an independent research field and is it going to be combined with other research disciplines?
– The elegant combination of signal processing, control, mechanics etc. is actually hidden in products. Is the value of control not seen anymore?
– The level of teaching automatic control has in general been of high quality. In the industry the graduated students are respected and have obtained good positions. How will the situation be in the future? How can these kinds of issues be "controlled"?
– Should the universities and universities of applied sciences specialize in teaching only some application areas of automation or keep the wide view?

- How is the financing of teaching automation going to develop in the near future?
- What must a graduating student from the automation field know and master now and in the future?
- What about the connections and networking by teachers, students and industrial representatives on a regular basis?

It is interesting to note that the "threats" towards education of Automation have not turned into reality, at least for the time being. Actually it is quite the opposite. For example in Aalto University the master program of *Control, Robotics and Automation* is the most popular in the School of Electrical Engineering. The industrial relevance is obvious and the alumni from the industry are extremely happy with the new employees graduated from this program. The educational target and objective would need to continue that way.

The purpose of this paper is to discuss the educational challenges in Automation and Automatic Control from the viewpoints of a. challenges of automation education for the society of future and b. digitalization of education as a new discipline of teaching an learning. The idea is to describe a framework for how to arrange studies and study entities in a new way, making distant teaching and learning possible as much as possible. Some items to be discussed in detail in the paper are

- Courses arranged such that the students are divided into small groups (2-3 students) working in continuous and close contact with the teacher. Group work, projects by computer using industrial software, seminars form the syllabus of the course. No exam.
- Digitalization project is described, which is ongoing in Aalto University, School of Electrical Engineering. Many short lecture sessions, which are recorded (Zoom, Panopto). Exercise sessions are also recorded and taught on-line for the students to participate and ask questions. Slack problems, homework problems, intermediate and final exam are included.
- Adaptive Learning and Teaching is an environment, where feedback from the students are continuously obtained and used for modifications in teaching sessions. This is an interesting but extremely difficult exercise. It is definitely something quite new in education, even though feedback is nowadays collected in all courses in the end phase of the course.
- Learning and Teaching Nuggets. This in an large international exercise, in which the Aalto Factory of the future is participating. The idea is to form a large number of learning entities (some of them "small", some "large") where the students can learn in a digital environment and get a certificate of his/her knowledge of that topic. Actually, this kind of certification of knowledge and skill is something that is on-going in the Automation field in Finland. It is again something new at least in the University level, and worth considering.

The paper discusses some of the above topics and tries to achieve one possible way for a framework for Automation studies in the future. In Chapter 2 traditional and modern teaching methods are discussed paying special attention to the change to digitalization in teaching and learning. Chapter 3 describes the student attitudes and experiences in distant learning. That topic is extremely important especially in respect to the continuing pandemia situation. Student well-being is a contemporary key concept, which is also a topic of great concern. The influence in teaching Automatic control are further discussed in Chapter 4 and conclusions are given in Chapter 5.

2 Digital era: past and present

2.1 Traditional teaching

The traditional teaching has mostly been based on lectures and exercise hours. That form of teaching is cost-effective, because - exaggerating a bit - it does not matter whether the number of students is very few or extremely large (several hundreds). Also, the burden to the teacher(s) is relatively light, the only big effort being the evaluation of exams. The weaknesses of this kind of teaching have been recognized, one reason being that the students remain quite passive during the studies. It has been reported that in some universities (not in the technical branch) students could graduate in master level "without speaking one word in any study session" during their whole study time. It is extraordinary that this old-fashioned way of teaching prevailed so long time. This may be partly because of the Academic ancient tradition and partly because teaching was earlier considered to be of minor importance in the university level. It is also true that many students appreciate the

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

mass lectures. The reason for that is that they look forward to university level teaching according to the ancient and respected tradition in Academia. Also, some students may also consider it to be more convenient remaining passive listeners during study sessions.

In addition to lectures and exercise sessions, design assignments, seminars and laboratory work were added in the study courses and curriculum. That extended the scope of studies already, but in practice not so much. At least in the technical studies in universities the seminars were few and mostly took place in the very end of studies. The home assignments were few, maybe one or two in the course, and the laboratory exercises also just a few or organized such that they were collected into one special course. Even today, special laboratory courses are often arranged in order to "take care" of practical experiments in studies, and the inclusion of laboratory work in the normal courses is not very common.

The main problems in the above study schema from pedagogic viewpoint were twofold. Firstly, the students were passive elements. They were expected to listen and learn, and self-activity was minor mostly through exams only. Secondly, the course evaluations were almost totally based on final exams or two intermediate exams. That distorted the studies during courses to focus only to the final parts of the course, and no continuous learning during a course took place. The author of this paper has heard from several students that "surely we do not really study the course until just before the exam" and "nobody uses more than two days to prepare for the exam". Assuming that these oral statements are true among most students, it is no wonder that the real learning output was not very good. Actually, the third pedagogic weakness of traditional teaching becomes obvious: When exams were in practice the most important way to evaluate students, surface learning just before the exam was often the result.

2.2 New ideas and methods

The new era of teaching and learning has brought pedagogical experts in the game. At least partly based on their recommendations the culture has been changing. Also, it should be pointed out that the importance of teaching is nowadays emphasized also in the university level much more than earlier. For example in Aalto university lecturer and professor nominations the teaching ability and expertise of the applicants is carefully evaluated by teaching portfolios, lecture demonstrations and interviews. Earlier in university level teaching was considered to have a minor role, research being the only thing that counts. That has and is changing rapidly, because the importance of good teaching for the game changers of the future is clearly seen as a solution to the need for specialists of current and tomorrow's technology.

The modern ideas are based on continuous teaching, learning and assessment. Some of the key methods are:

– Decrease the number of (mass) lectures.
– Organize the courses as lectures/short lectures, exercises in groups, assignments in groups, group seminars, laboratory work included.
– Divide teaching to happen partly in classrooms or in negotiation facilities, partly digitally and possibly recorded.
– Arrange skill checks, assignments and exams during the duration of the whole course, not only in the end phase.
– Use *flipped classrooms*, discussion and group work
– Arrange project work in co-operation with the industry.

There are a couple of main points to be considered in the development of teaching by the above ideas. Firstly, the teaching can be separated by "classroom" kind of teaching and distant teaching. Secondly, the way to use digital solutions must be clearly investigated. Working together in small groups has been proven to be a proper way towards continuous learning during the course when compared to mass lectures. To keep up with the students' interest problem solving sessions, discussions and presentations is groups form an effective way of pushing the students to be active and productive all the time. Yet, that is challenging also: it forces students to work and participate, even though some students might not like this.

The digital solutions in teaching can be seen as a form and possibility towards distant studying. Group sessions can be arranged as virtual meetings, lectures (or short lectures) can be arranged that way also and recorded for the students to follow afterwards also. Skill checks for example before or after each session can be made and evaluated digitally (there are platforms for that, and many universities use effective study portals already). Exercise hours and home assignments can be given digitally. Instruction can be given on-line for example by arranging meetings for the students so that they can ask questions directly and/or ask questions in writing by on-line applications (e.g. Zoom, Slack, Tele-

ISBN: 13 978-952-5183-59-7

gram etc.). The best ways to organize this kind of multi-instruction in the most effective way is currently investigated quite much in Finland and elsewhere also. In Aalto university there have been several reports on the main problem, which is somehow astonishing: Nobody seems to know how exercise hours should be arranged in distant learning such that the students would participate actively.

Assessing the students is another important topic. There is a trend of avoiding (or at least decreasing the number of) "old-fashioned" mass lectures and full exams and go towards more modern methods. As an example, let us consider a basic course of Control Engineering in Aalto university. This is a mass course of about 250 students, which makes it difficult, or even impossible, to introduce teaching sessions in small groups (digitally that might be possible, in classroom teaching hardly). The course contains lectures, lecture Quizzes (skill checks), exercise hours, home assignments, two intermediate exams and alternatively one full exam. Everything happens in distant learning mode. Lectures consist of both longer (traditional) lectures and small information lectures on a special topic (e.g. 15 minutes). Some of these are lectured normally in Zoom, but everything is recorded and saved in the study portal (MyCourses) so that the students can watch the material later also. In connection to every lecture there is a skill check (Quiz), which is done and evaluated automatically in the portal. The skill checks are mostly True/False and Multiple choice problems related to the previous and the next lecture. Exercise hours are given by Zoom sessions and Slack, so that the students can solve exercise problems and also ask help either orally or by writing. Also, help for home assignments can be received this way. Home assignments (altogether six) are more challenging problems, including small controller design problems to be done and verified by simulation. They are the key material of the course in order to develop the students' ability to problem solving instead of just solving simple exercises or answering questions by writing. The weight of the home assignments in course evaluation is 40 percent, Quizzes 10 percent and exams 50 percent.

The intermediate exams or full exam are done at home. The questions appear in the study portal in a given time for three and a half hours (three hours for solving the problems and a half an hour for submitting the solutions in the portal), The students can use a text editor with the possibility to produce equations, figures etc. For example Word and LaTeX are possibilities in that. The exam is "open book", so the students can use course material, pocket calculators and simulation software. However, with their signature they assure that they do not use other alternatives to get information, nor be in any contact with another person during the exam.

The modern structure of the course becomes apparent by looking at the course structure: the students must be active through the whole course. By only exams they can pass the course only theoretically and with a bad grade (weight 50 per cent). In home assignments (weight 40 per cent) group work is allowed, actually recommended and encouraged, but not organized by the teaching staff.

The exams in the course are only for checking that each student individually has learned the basic and most important concepts and methods taught in the course. Since group work is encouraged otherwise, the exam part is also considered important, to exclude the possibility of "free riders". Also, in the course the students must really solve problems and do design, which they must show correct by simulations. There have been some suggestions by the pedagogic experts that the studies should more and more be based on discussions by groups and writing essays. The author of this paper does not support these kinds of ideas in exact natural sciences and technological science.

In the new study practices there is the possibility of "free riding" and cheating. There are some opinions that even 10 percent of the students might cheat, if there is a concrete and easy way to do so. That is a difficult issue. For example, to superintend the students in an exam done from home might bring difficult law issues into the game.

It is interesting to note how much the possibilities for digitalization of education has progressed in 10-15 years. Then, the first web courses were created for teaching of Control Engineering in Helsinki University of Technology (now Aalto University). The courses *Basic Mathematics in Control*, *Analog Control* and *Digital Control* contained full text materials developed specially for the web course, interactive Quizzes and special exercise sets with detailed solutions. One of these courses used the *Stack* engine, by which problems could be varied considerably more than just changing some parameter values. Even though the named courses were not network courses in the true meaning (they were only used as extra study material instead of forming separate course entities) they were used in several universities in Finland. However, from current point of view they were rather clumsy and constructed by old-fashioned technical solutions.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

In Aalto university the general study portal *My-Courses* is used. It is a Moodle-based solution, which is relatively easy to use for course information storage (lecture slides, exercises with solutions, home assignments, recordings) and student information platform. The problem is that it does not really support on-line teaching and learning. Therefore, additional platforms like Zoom, Teams, Slack) have been used to make on-line interaction with the teachers and students possible. This works, but the problem is the use of several contact platforms in one course, which might cause some confusion and difficulties to the students.

There are other platforms to generate on-line courses, which have been tested. For example the *TIM* platform, developed by the University of Jyväskylä has been used to generate full course learning environments including study materials, interactive problems, exams etc. The use of a mathematics engine makes it possible to construct problems that vary considerably at each exam or each homework for example. In Aalto University the course *Signals and Systems* has been a testbench for TIM, and the experiences have been very good. The course which earlier got a lot of criticism for its difficulty is now being evaluated as one of the best courses. The problem is the same as with other similar kinds of platforms: Developing the course with materials and problems is very slow, needs expertise and anyway involves a considerable amount of work. In addition, the fact that the system is not in the university's own hands, makes it fragile and questionable to use in a large setting.

Other initiatives also exist. For example, in Aalto University School of Electrical Engineering there is an on-going project for the development of digital courses systematically and taking the pedagogical aspects into account from the very beginning. After preliminary study and literature survey with similar projects abroad a few courses have been selected for development. The work is on-going and there are no reports on the results yet. It is however possible that the project will constitute a framework for systematic development of digital courses in Aalto University School of Electrical Engineering in the near future.

In summary, it seems that there exist a lot of possibilities for modern teaching and learning with a multitude of platforms available. It is believed that distant learning will continue at least to an extent after the covid pandemia also. The main questions then are:

– What platforms are to be used for on-line interactive teaching?
– How will the course development be done by a reasonable work by teachers?
– How do the teachers feel about this?
– How do the students feel about this?

3 Student well-being

The last question in the end of the previous section is of utmost importance. The student well-being has been identified to be a critical issue and target for concern even before the covid pandemia. That also reflects the idea that times in the digital era are changing. Much earlier the attitude at least in universities seemed to be something like "let the students swim and see which ones survive". Nowadays the student well-being is taken very seriously, and questionnaires and plans for improving it are constantly being made.

In Aalto University School of Electrical Engineering an All-Well project has been established. In May 2020 a questionnaire was sent to 1744 students regarding their opinions and feelings during the corona age. There were 337 responses (19,3 per cent), out of which 28 students asked to be contacted by LES (Learning Services). They were all contacted within one week from their reply. 86 per cent of them wanted to discuss their studies, 36 per cent wanted to discuss their coping in general. The answers were diverse in many aspects. Most students were happy about the quality of remote teaching. For example, they emphasized the possibility to watch the teaching videos afterwards and many times. Somewhat surprisingly, many respondents wished the remote teaching to continue after the pandemia also. However, some reported problems related to loneliness, difficulties in motivation, problems because group work was difficult in the remote phase etc. Many students wished to have more support from the teachers, for example in the use for the technical equipment used in remote courses. Some wanted to have special facilities for meeting workgroups at least sometimes, for example in the midterm.

One of the first concrete means to deal with the above issues has been to establish a support group for first year students. The pandemia time has been very difficult for these students, because they have not been able to meet each other and make friends with older students. In short: grouping has been almost impossible to them. The support groups are intended exactly to this end to help with all issues from social relationships to studies in different courses.

The above results seem to indicate that the university students might not be so badly off in the corona age as some other and maybe younger students at schools.

However, without proper research results available that is difficult to know for sure. For example, it is not known what the above mentioned 36 percent of respondents wanted to discuss with LES concerning their welfare and coping.

4 Influence on Automation education

The Covid 19 era has shown the way to a considerable increase in remote working, remote teaching and remote learning. That has immediate consequences. Companies are considering to continue with greater amount of remote working after the pandemia than nobody could imagine before. Universities have, somewhat astonishingly, similar kinds of ideas concerning administrative work, research and teaching. It is anticipated that in the near future part of university education takes place in the campus and part as distant learning. The experiences from the past one year have shown a direct way to that.

In teaching Automation and other disciplines as well the above must be regarded as strength, not weakness. Modern digital teaching and learning environments have been and are being developed for modern study systems. Key concepts are to establish a firm theoretical background, do group work, projects and project work, establishing co-operation with the industry and internationalisation. For example, in Chapter 1 the teaching nuggets have been mentioned. Special "Factories" are being developed, which are aimed at learning by practice and preparing for the technology solutions of the future. The industry is involved in these activities. Teaching nuggets are special learning units developed in international co-operation. That can be reflected to similar ideas which have been developed in Finland by some Universities of Applied Science: the systematic certification of well documented skills, by which the students can show their expertise. All this work is not ready, and it is hard to say what the reality in teaching in the time period 2-10 years from now on will be.

In general the teaching of Automation and Automatic Control seems not to be in the danger zone. They have so much applications in so many application areas that embedding them into some other disciplines does not seem to be a realistic option. The same thing has been observed by the new generation of students. Very many of them are confident that Automation gives a possibility for interesting work in the years to come. Therefore they apply for Automation studies. At least from Aalto University's perspective their number and quality are worth celebrating.

5 Conclusion

In this work some experiences and ideas of digital teaching and learning in Automation and other technological fields have been elaborated. The inspiration has been the one year long pandemia age, during which an abrupt change into remote teaching took place. Ideas for the future from both practical and pedagogical perspectives were given, but nobody really knows exactly, how the digital teaching era will develop in the next 5-10 years. The report and ideas presented in this text are based on the author's earlier research in the topic, and his experiences over a long teaching history, where especially the last one year plays a significant role.

References

[1] K. Zenger. Control engineering, system theory and mathematics: the teacher's challenge. *European Journal of Engineering Education*, 32(6):687–694, 12 2007. https://doi.org/10.1080/03043790701520719.

[2] Kai Zenger. Challenges and new directions in control engineering education. In *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation*, pages 819–824. Linköping University Electronic Press, ISBN: 978-91-7685-399-3., 2016.

Anna Hakala* and Aura Rapatti

Data-driven approach to stabilizing an effluent treatment plant (ETP)

Abstract: An industrial manufacturer experienced problems with a closed-loop effluent treatment plant (ETP). Unexpected ETP behavior caused production losses as the processed wastewater did not meet the environmental quality standards. Process-based COD modelling and the Valmet Dynamic Centerline Advisor were utilized along with operation analysis, in a Valmet Industrial Internet (VII) Data Discovery, which revealed that ETP operation conditions had not been ideal for biomass growth. Based on Data Discovery results, the customer revised their water balance management system to compensate for the cyclic nature of their production process. The actions taken proved to be effective and the ETP plant is no longer a bottleneck for their production.

Keywords: Effluent treatment, Data Discovery, data-driven problem solving, process optimization

***Corresponding Author: Anna Hakala:** Valmet Automation Oy, E-mail: anna.hakala@valmet.com

Second Author: Aura Rapatti: Valmet Automation Oy, E-mail: aura.rapatti@valmet.com

1 Background

In order to minimize use of fresh water in industrial processes, closed-loop effluent treatment plants (ETPs) can be used. As a byproduct, ETPs can produce biogas, which can then be used for energy production. [1]

A Valmet customer was having problems with a closed-loop ETP operation. Unexpected ETP behavior was causing major production losses as the processed wastewater did not meet environmental quality standards and the ETP load needed to be reduced from time to time by shutting down production.

The customer was using an anaerobic upflow reactor where anaerobic microorganisms are utilized to degrade carbohydrates in the wastewater into methane, thus reducing the chemical oxygen demand (COD), a measure of degradable organic waste in wastewater. [1] The industrial manufacturing process determined influent properties, and the mill personnel had difficulties adjusting the reactor operation in order to reach final effluent targets.

VII Data Discovery [2] provided a data-driven approach to finding root causes for the problems and identifying improvement potential. In addition to data analytics, process expertise was utilized to enrich analytics and create value for the customer through meaningful and cost-effective solution proposals based on identified improvement potential.

2 Aims

The aim of the Data Discovery study was to combine and analyze laboratory and process data of the whole industrial complex to reveal reasons for the ETP problems. Uncovering hidden correlations and causation between the customer's production process and the ETP operation enables stable operation as well as taking preventive actions when necessary.

3 Materials and methods

Existing tools from the Valmet Industrial Internet portfolio were utilized for the following research questions:

- Can a simple root cause explain the unexpected behavior?
- What is affecting the amount of incoming COD load?
- Can an unexpected root cause be identified that correlates with the ETP behavior?

3.1. Data

Data from over 8000 continuous process tags on a 10-minute sampling frequency were extracted from customer's automation system. Laboratory data of operator-performed measurements from the ETP were also received, containing mostly chemical measurements about the wastewater composition. Approximately 1 year of operation was covered by the data, and the data was prepared for analysis with an internal standard VII toolbox.

3.2. Operation analysis

Operation condition analysis was conducted with basic programming tools for data visualization and simple calculations of derived measures. Actual operation was compared to recommendations in the manuals, and a literature review was conducted to verify findings and their importance. Internal ETP expertise was also utilized to review the results.

3.3. Process-based modelling of COD

COD modelling utilized predictive modelling tools from the VII toolbox. Due to lacking delay information from process measurements, the first draft models preferred output measurements from the ETP. This effect was compensated by customizing the model to only include measurements taking place before the ETP.

The VII toolbox also contains a tool for analyzing process delays, but a detailed study of process delay effects was excluded from the Data Discovery scope. The missing delay information was considered when analyzing the results.

3.4. Valmet Dynamic Centerline Advisor (DCA)

The Dynamic Centerline Advisor is a VII process optimization tool for finding the best available operation point. Performance indicators and their targets, defined with the customer's process engineer, were used to formulate a Centerline Index which describes how well the defined targets are met. This behavior index is used for in-depth analysis of the underlying root causes based on process variables available in the process data.

The performance indicators selected for Centerline Indexing included wastewater quality measures at inlet and outlet, reactor operating conditions such as temperature, pH and upflow velocity, and reactor health indicators.

4 Results

The study consisted of two phases:

- Initial analytics to assess the data quality and find the most interesting research tracks
- Advanced analytics to dive deeper into the defined problems and their causes.

Initial analytics results were presented to Valmet's ETP expert as a standard procedure to validate the initial findings and to define objectives for advanced analytics. Advanced analytics results were again confirmed with the ETP expert before presenting to the customer. The main analytical results are presented.

4.1. Operation analysis

The Data Discovery revealed that the ETP operation conditions had not been ideal for carbohydrate degradation and biomass growth, as stable operation conditions are a prerequisite for stable performance. COD concentration at the reactor inlet varied due to the manufacturing process and its disturbances, furthermore, the daily COD load was not stable as shown in Figure 1. The microorganisms degrading the carbohydrates of the wastewater do not work ideally in such conditions.

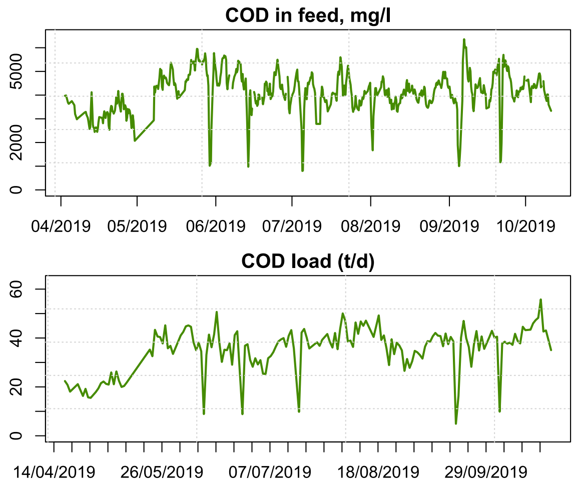


Figure 1. COD concentration at the reactor inlet and daily load of the reactor. Optimally these would be stable.

In addition to the COD instability, ETP reactor inflow and thus the upflow velocity inside the reactor had been significantly below the suggested target, as shown in Figure 2.

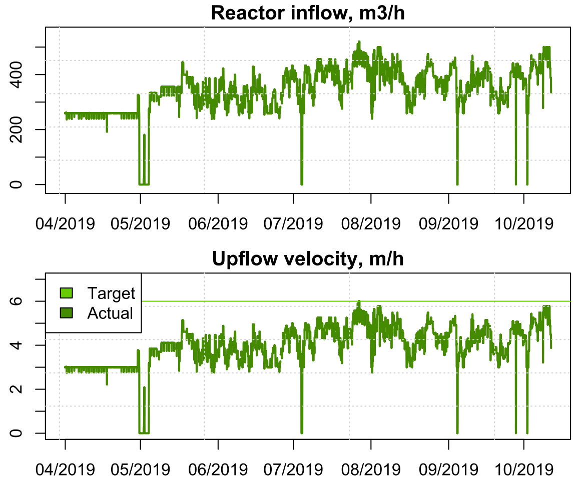


Figure 2. Insufficient reactor inflow results in a decrease of upflow velocity inside the reactor. Continuous upflow of the fluid is needed to make sure the excess calcium carbonate is flushed out.

Together with the high calcium concentration, the low upflow velocity resulted in an increased biomass crystallization rate and biomass flushing out.

4.2. Process-based modelling of COD

Predictive modelling of the COD concentration resulting from the production process was promising. R^2 values were around 70 %, meaning that the incoming COD level could be predicted based on the automation system data depicting the manufacturing process. However, according to the study, the prediction accuracy would most probably improve by taking the process delays into account, as it would balance the impact of chained sub-processes.

One notable result from COD modelling was that the most important features selected by the model were related to the pH of the manufacturing process. Low pH originating from the process rather than the pre-acidification tank increases dissolution of calcium, and higher pH in the anaerobic reactor can result in calcium carbonate precipitation.

4.3. DCA results

Dynamic Centerline Advisor implementation and analysis were successful and the monitoring capability was proven in the study. Low reactor performance time periods were mostly connected to low inflow and COD values at the inlet of the reactor, suggesting that the load of the reactor was not stable over time.

Identified correlations between the Centerline Index and process variables verified the findings from operation analysis and COD modelling. It was also noted that introducing delays would probably add analytical value to the application, as the reactor health indicators react slowly to changes in reactor conditions.

5 Conclusions

The Data Discovery provided a data-driven, analytical approach to identifying and fixing problems the customer was facing with their effluent treatment plant. The Data Discovery revealed that the problems arose from unstable operation of the reactor, which posed practical difficulties due to the biological nature of the degradation process.

Operation analysis and COD modelling revealed that the operating conditions were not ideal for reactor performance. The COD load of the reactor was not stable and the reactor upflow velocity was too low to flush out the excessive calcium carbonate present in the reactor. These findings were also confirmed with the Dynamic Centerline Advisor analysis. Further research on process delays with the Dynamic Delay Calculator was also suggested.

Based on the Data Discovery, actions for stabilizing the COD load and reducing the wastewater calcium concentration were recommended. Based on the recommendations the customer revised their water balance management system to compensate for the cyclic nature of their production process.

The actions taken by the customer have proven to be effective, and the ETP plant is no longer a production bottleneck. The customer learned and implemented new ways to operate the whole plant in a more stable and predictable manner, and the major production losses experienced in the past can now be avoided in the future.

6 References

[1] Emis Vito: Anaerobic Biological Wastewater Treatment. Available: https://emis.vito.be/en/node/22498, cited 8.3.2021

[2] Valmet: VII Data Discovery. Available: https://www.valmet.com/automation/industrial-internet/vii-data-discovery, cited 9.3.2021

Gerardo Santillán Martínez*, Jari Lappalainen, Antti Jäänvirta, Tommi Karhela

Cloud-based LCA dashboard for improved sustainability in the process industry

Abstract: Life cycle assessment (LCA) is an end-to-end analysis to calculate the environmental impact associated with a given product and its production process. These analyses are commonly carried out once during the design phase and on a yearly basis during the production process. LCA results are required to comply with environmental guidelines. During the last decade, ever-increasing environmental regulations and customer awareness have forced owners/operators to perform LCA studies more frequently. Although proposed approaches, such as Online LCA, can reduce the effort to periodically obtain LCA results, system integration remains laborious. In practice, data collection required for LCA is still carried out manually due to the slow industrial adoption of software architectures for easier data integration and collection. This paper presents an online LCA solution for industrial processes that leverages both, Online LCA and a cloud software architecture for continuously calculating and representing up-to-date environmental assessment and sustainability-related information. A prototype was implemented in an industrial pulp production process. The initial results show that sustainability professionals can significantly benefit from the continuously updated LCA results.

Keywords: Life Cycle Assessment, LCA, Online LCA, sustainability, cloud computing

***Corresponding Author: Gerardo Santillán Martínez:** Semantum Oy., E-mail: gerardo.santillan@semantum.fi

Jari Lappalainen: Semantum Oy, E-mail: jari.lappalainen@semantum.fi

Antti Jaanvirta: Semantum Oy, E-mail: antti.jaanvirta@semantum.fi

Tommi Karhela: Semantum Oy, E-mail: tommi.karhela@semantum.fi

1 Introduction

Life Cycle Assessment (LCA, also known as Life Cycle Analysis) is a “cradle-to-grave” analysis of the cumulative potential environmental costs associated with a given product and its production process [1]. LCA is an end-to-end analysis that considers raw materials, transportation, production processes, usage, and disposal of a product. LCA models are used to estimate potential direct and indirect environmental impacts associated with the production and use of a certain product [2]. Results are commonly used for supporting decisions of policymakers and environmental authorities. They are also used by purchasing departments in business-to-business (B2B) interactions. Additionally, LCA results are used by sustainability and brand managers as a sales argument to answer the pressure coming from consumers, who in turn are increasingly more aware and concerned of their environmental footprint.

LCA analyses are commonly carried out only once during the design of the production facility or of the production process. This assessment is required in the early design phase to obtain the environmental authorities’ approval for project implementation and eventual operation of the facility or production process. Recently, due to tighter targets on the reduction of CO2 emissions, policymakers and environmental authorities have started requesting results of LCA analyses also on a yearly basis. This is required to assess and monitor the environmental performance of the production process also during its operation. Therefore, LCA analyses must be carried out yearly, based on actual production data as well as on updated production downstream and upstream information.

Recent initiatives such as Product Environmental Footprint (PEF) [3] from the European Commission will in the future turn into compulsory legislation. Consequently, LCA results of production processes would be required more frequently than once a year. The LCA analysis is a laborious task to perform, and this is usually done manually. The related data collection is time-consuming due to many information sources and lack of integration between the information systems that contain the required up-to-date data. Moreover, LCA results are typically available only to LCA experts, hindering wider environmental impact awareness of the production.

Online LCA methods could be applied to increase the impact of LCA analyses [4][5]. In this approach, the LCA model is directly interfaced with the production process' information systems. Consequently, data collection and analysis are continuously carried out based on the production process data, such as raw material flows, chemical additions, and energy consumption. However, it is important to note to whom the LCA information mostly matters within the plant and company personnel. Today, the sustainability indicators are not necessarily relevant in the day-to-day operation of the process. Furthermore, cloud technology could be used to provide wide access to the results, and to ease the implementation and maintenance tasks related to the connection of the LCA models with the information systems and process databases. Additionally, cloud-based systems enable the development of collaborative applications in which data can be easily accessed, visualized, and communicated by different stakeholders throughout different organizational levels.

This paper presents an LCA solution for industrial processes that leverages on the Online LCA approach and cloud-based software architecture for continuously calculating and representing up-to-date environmental assessment information. This system relies on an intuitive user interface (UI) to present the calculated environmental key performance indicators (KPIs), as well as economical and other production KPIs when needed for the results interpretation. These KPIs are selected and presented per the end users' interests and needs. The presented system is a decentralized software solution that can be collaboratively used for decision support, not only by LCA experts, but by other stakeholders such as production managers and company executives. It enables increased and continuous awareness of the current environmental footprint of the production, and thus also helps to understand the changes on the road to more sustainable production.

In this work, an industrial pulping process was used to build a prototype implementation of the described system. The LCA model of the production process was developed in the LCA software SULCA [6]. In the proposed system, the LCA model is deployed on the cloud as a Functional Mock-up Unit [7] and then interfaced with the InfluxDB [8] time-series database. The pulp production process information is collected from the process monitoring systems. Finally, a dashboard-based user interface was developed to present the calculated data using the InfluxDB visualization tools.

2 RELATED WORK

LCA is an analysis of the environmental impacts associated with a given product and its production process that considers all the resource requirements, as well as the different materials and energy flows accounted for an inventory process [1]. LCA also involves the analysis of Life Cycle Inventory (LCI), such as in [1], to provide quantitative descriptions of the energy/mass efficiencies of each unit process steps, from taking raw materials as inputs until the final steps for manufacturing and delivering the product. LCA also includes, life cycle energy analysis (LCEA) and life cycle cost analysis (LCCA) studies such as in [9] [10], which present analysis of the environmental costs and impacts of buildings and building related industry and analysis of mechanical and chemical processes to produce energy from renewable sources. The requirements and reporting information that should be provided by LCA analysis is gathered in international standards, such as the ISO 14044:2006 [11].

Recent efforts to enable continuous calculation and easy accessibility of LCA KPIs, have resulted in the development of the Online LCA approach. In this method, direct and persistent communication between the LCA model and the physical system enable continuous LCA model computation. In the existing Online LCA implementations [4], [5], the LCA system is connected directly to the automation system of the production process. Therefore, LCA calculations are based on real-time information collected from the control system. As a result of this connection, after been properly configured, the task of gathering the input information for the analyses is avoided. LCA and environmental KPIs of different periods of time are continuously obtained. So, LCA trends can be collected from changes between different production shifts – change of the chemical mix, different raw materials, seasonal change, etc. and used for different analyses just like any other process variables.

Although Online LCA allows continuous and efficient calculation of LCA KPIs, system integration between the life cycle assessment system and the actual production system is still complex and thus time-consuming. At the same time, security conditions to establish direct connection with the automation system may vary depending on the production process. Moreover, if the online LCA system is located close to the control application, visualization of the LCA KPIs is usually limited to the personnel at the production site. This hinders decision making based on the LCA results at other locations. This becomes more critical when decisions based on LCA results must be made considering not only a single facility but a fleet of production facilities. Cloud technology could be applied to tackle system integration issues and to provide easy access to the LCA information for any authorized person in the organization, independent of her/his

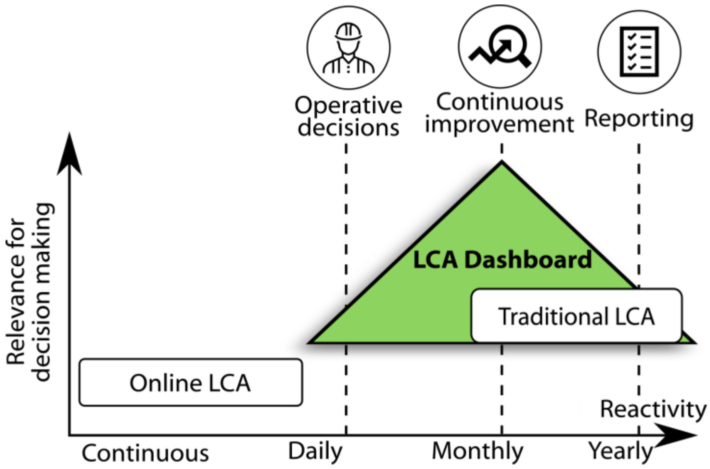


Fig. 1. LCA Dashboard Reactivity versus Relevance of decision making and its comparison with other LCA approaches. Optimal reactivity of LCA Dashboard can vary according to the end-user needs.

location.

3 Cloud-based LCA dashboard

3.1 LCA Dashboard concept

LCA Dashboard is a concept linked to the Online LCA approach, as both provide means for developing continuous environmental awareness of the production. However, LCA Dashboard differs from Online LCA, as this system is not directly connected to the automation system of the corresponding production process. Instead, it interfaces with the monitoring and/or information application(s) of the production system. Additionally, unlike existing Online LCA implementations, LCA Dashboard results are not only targeted to the personnel in the production. This is because LCA information is not as relevant within plant operators as it is within managerial and corporate level decision makers. In fact, sustainability indicators are not, at least for the time being, particularly relevant in daily process operation. For this reason, LCA Dashboard employs a cloud-based software architecture that facilitates visualization of the LCA results from different production facilities, at different sites. This information is available from a single application that can be accessed from anywhere. LCA Dashboard promotes organization-wide environmental awareness based on continuous KPI calculation from up-to-date production and inventory information.

Fig. 1 outlines different LCA approaches in terms of the frequency to calculate KPIs over time (reactivity) and how relevant their results are for decision making. Traditional LCA is performed on a yearly basis to comply with environmental guidelines. On the other hand, Online LCA is continuously calculating KPIs and persistently presenting its results to the process operators. In contrast, LCA Dashboard is a middle ground solution in which KPIs are obtained and displayed to decisions makers across the production process and its organization. The optimal time resolution for the KPI calculation varies according to the end-users within the organization. In some cases, more regular LCA results are required to, for example, take control decisions to reduce emissions. In other cases, medium-term calculations might be needed to take decisions related to purchase and inventory or to communicate such results to customers or other stakeholders. Similarly, long-term LCA Dashboard results can be used for reporting to environmental authorities as requested in environmental regulations.

3.2 LCA Dashboard cloud-based architecture

The software architecture of the system proposed in this work is presented in ***Fig. 2***. The approach is named LCA Dashboard, as the system results are presented using a dashboard view.

The main component is the so-called Dashboard coordinator. This component comprises the LCA model, as well as coordination mechanisms to retrieve and process the relevant data required for performing the LCA; coordinate the LCA model simulation; and to provide data back to the database for its visualization. The LCA model (green in Fig 2) calculates the LCA KPIs. Similarly, other sustainability-related data processing can be carried out by the Dashboard coordinator. The coordinator runs the LCA model based on the Functional Mock-Up interface (FMI) [7]. FMI is a standard for co-simulation of heterogeneous

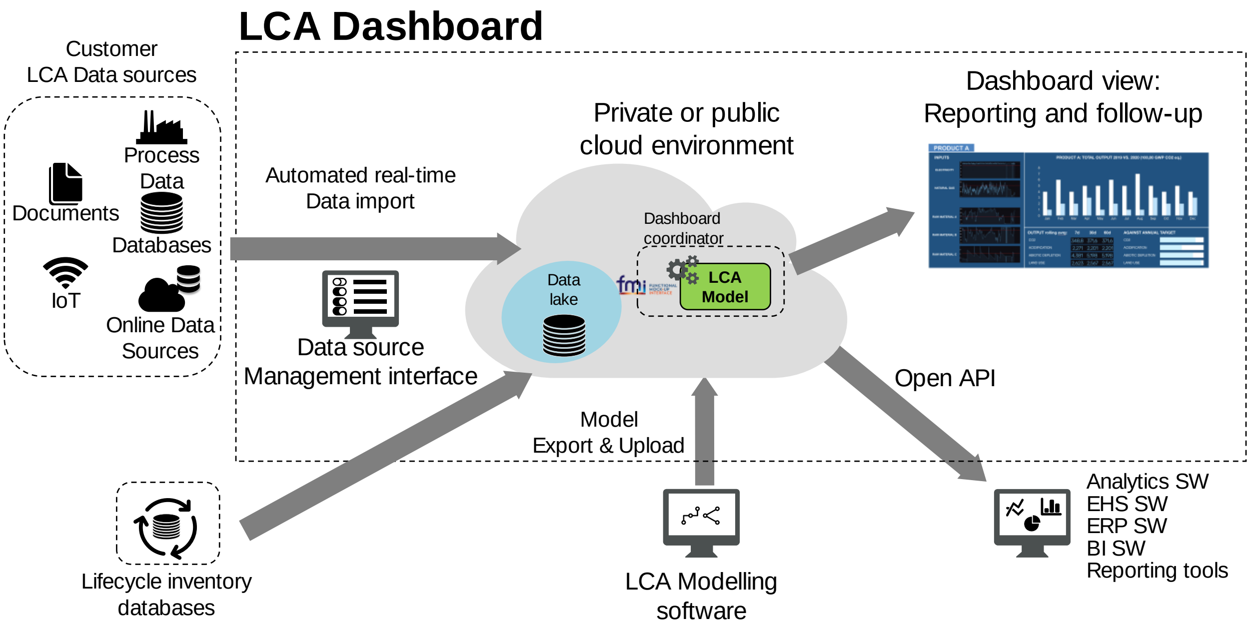


Fig. 2. Proposed cloud-based Online LCA system architecture

simulation models. It is a widely used standard in the research field and its adoption is growing also in the industry [12][13]. Using the FMI allows LCA modeling experts to interface also existing LCA models developed with tools that supports this standard. At the same time, FMI reduces implementation effort required to integrate LCA models with other components of the architecture.

The LCA Dashboard architecture also includes a data source management interface, required to handle the automated, real-time data import from different data sources. These sources, which vary according to the targeted production system and the end-users‘ needs, include a combination of documents, process historical data, process databases, IoT platforms, and monitoring systems as well as of other online data sources. The Dashboard coordinator oversees the import, storage, and management of data through its source management interface. Imported data is stored in the Data Lake system. The source management interface also communicates with the lifecycle inventory databases needed for the LCA calculation. These databases provide reliable and up-to-date inventory data for a large variety of products. This is required to reduce effort for manually gathering this data.

The LCA Dashboard results mostly focus on visualization of the LCA analyses. However, other sustainability KPIs can be seamlessly calculated depending on the needs and data collected. Examples of these sustainability KPIs include recycling rate, different emissions to air or wastewater, as well as energy and water consumption. Both, LCA analyses results as well as other sustainability KPIs are shown using dashboard views. The reporting view of LCA Dashboard also includes options for visualization of follow-up actions based on the results obtained by the Dashboard coordinator. The dashboard views are configured based on the built-in visualization tools included in the cloud development environment. Additionally, the LCA Dashboards includes an external application programming interface (API) to which external software tools can connect. Examples of these external software include software tools such as data analytics, environmental health, and service (EHS); enterprise resource planning (ERP); business intelligence (BI); as well as other reporting tools. LCA Dashboard is a cloud service that can be deployed privately on the end-user’s own IT systems. Therefore, although it is a cloud system that can be accessed in different locations, it can only be accessed by authorized personnel.

4 Pilot implementation

In this work, an industrial pulping process was used to build a prototype implementation of the described system. The pilot implementation addressed kraft pulp production. In kraft pulping, wood chips are converted into cellulose fibers that is further used for producing different paper and paperboard grades. The process entails multiple steps, treatments by mechanical and chemical means. Besides wood, the process requires

ISBN: 13 978-952-5183-59-7

Table 1. LCA Dashboard KPIs and their categorization

CATEGORY	KPI
LCA & Sustainability	Global warming potential
	Human toxicity
	Marine aquatic ecotoxicity
	Ozone layer depletion
	Photochemical oxidation
	Terrestrial toxicity
	Terrestrial oxidation
	Ecosystem ozone formation
	Particulate matter formation
Cumulative Energy demand	Biomass demand
	Fossil fuel demand
	Nuclear energy demand
	Renewable energy demand
Production & efficiency	Capacity utilization
	Production rate
	Raw material usage/ Ton of prod.
	Water usage/Ton of prod.
	Energy usage/Ton of prod.

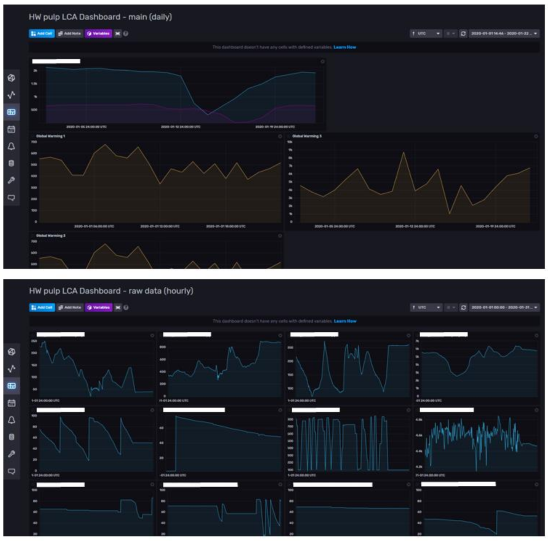


Fig. 3. LCA Dashboard user interface

chemicals, for example, in the cooking and bleaching subprocesses. The need for chemicals is significantly reduced by chemical recovery, which also produces the electrical energy needed and a surplus.

The LCA model of the mill was developed in the SULCA LCA software prior this project. SULCA is a tool for environmental, carbon and water footprints calculation. SULCA has been used to carry out LCA analyses and to support the evaluation of sustainability and environmental performance of various products, processes, and other systems [4],[14]. The existing LCA model of the of the pulping process was exported as an FMU. The FMU was deployed on the LCA Dashboard cloud installation and then interfaced with the InfluxDB database.

The described system was implemented on a cloud environment and connected to the process using the data source management import. The LCA Data Lake was developed utilizing InfluxDB database. The Dashboard coordinator interfaces with the LCA model using the functional mock-up unit (FMU) of the FMI standard for co-simulation. The LCA model FMU is a functional component that includes the model structure as well as the solver and methods for independently calculating the LCA KPIs. The Dashboard coordinator retrieves data required for the LCA analysis from the InfluxDB database. Before the data can be used for the LCA calculations, they typically require some pre-processing. This is required, as the raw production data may need conversion due to the measurement units or may have inconsistencies such as clearly faulty values; differences on the data time stamps; or irregular time sampling.

In the case presented in this work, the production-related information is collected once per day. Consequently, LCA and other KPIs are calculated and updated also daily. Once the pre-processed data is collected and after performing the required calculations, the Dashboard coordinator provides its results back to the database. These results are then visualized on the Dashboard and reporting view. The results are mainly LCA and sustainability KPIs. Furthermore, since it is also able to collect other important production data, LCA Dashboard is also able to calculate other economic, production and efficiency related KPIs of the process. This is particularly useful, not only for continuous environmental awareness, but also for supporting decision making of sustainability engineers, production managers and company executives. ***Table 1*** shows examples of the variables calculated and displayed on LCA Dashboard, and ***Fig. 3*** shows some source data and KPIs displayed on the LCA Dashboard view.

LCA Dashboard uses various process information sources to perform the LCA calculations. Therefore, it interfaces and collects information from the monitoring system of the pulp production process. Additional information is obtained from the LCA inventory database. LCA dashboard does not connect directly to the control application. Connection to the monitoring system is a more secure way of obtaining data required for the LCA analyses. This, as well as the easy access to the sustainability results, can potentially increase industrial adoption of this method, compared to other Online LCA implementations. Production-related information is collected daily. Consequently, LCA and other KPIs are calculated and updated also daily. ***Fig. 3*** shows the KPIs calculated displayed on the LCA

Dashboard view.

5 Conclusions

LCA assesses environmental impacts (e.g., carbon footprint) of product, process, or service "from cradle to grave". LCA is gaining popularity, yet the speed is limited by its scientific and laborious nature. This paper presented a novel concept called LCA Dashboard to contribute to the utilization of the LCA method in the process industry. The LCA Dashboard concept, here demonstrated with a prototype implementation, is based on a cloud software architecture. The proposed LCA Dashboard system aims to fill the gap between traditional LCA and Online LCA. It persistently connects the LCA model with the corresponding production process, similarly as in the Online LCA examples available in literature. In other words, it calculates and represents up-to-date environmental assessment information of the targeted production process. Furthermore, the proposed solution leverages on a cloud architecture that integrates in a single application, different mechanism for data collection; for LCA calculation; as well as for reporting calculated KPIs. This application can be simply accessed using a web browser by decision makers across different levels of the organization.

In the prototype implementation, the LCA model had been developed with the SULCA software. SULCA features FMU export for the model. Thus, the existing LCA model of the target production plant could be exported into a software package that is run in the cloud as part of the LCA Dashboard solution. Besides the model, the FMU includes stand-alone solver functions and FMI interfaces for the input/output communication, making the LCA calculations seamless. The FMU that features the LCA model could have, as well, been exported from another LCA tool that supports the FMI standard.

The cloud architecture of the presented system relies on the InfluxDB database. This solution provides a toolkit for developing dashboard user interfaces that were used for visualization of the calculated LCA and sustainability related KPIs. Although the InfluxDB's dashboard development toolkit is a comprehensive set of tools, other front end visualization tools can, as well, be used to intuitively present the sustainability information.

The LCA Dashboard concept aims to make companies' sustainability experts and management more aware of the current environmental footprint, but especially of the connection between the production facts and the sustainability impact in a longer run. We believe that making easy access to the relevant sustainability information is a key factor towards more sustainable industry.

Current trends of increasing consumers' environmental awareness, along with investors' growing willingness to reward best strategic sustainability performers [15], will soon force industrial companies to rely more on LCA. At the same time, consumers have also started trusting LCA results to understand and compare detailed sustainability figures of the product in their hands. With LCA being the only feasible method to focus on product level sustainability metrics, and upcoming EU PEF legislation banning marketing of products as 'green' without LCA proof, it is clear that companies should deploy LCA Dashboard –like solutions to start building awareness throughout their organizations. While LCA results are not always relevant for stakeholders on all organization levels, they should be easily available and changes easy to monitor.

References

[1] David M. Prett, Carlos E. García, and Brian L. Ramaker, *The Second Shell Process Control Workshop* , 1st Edition. 1989.

[2] O. A. Asbjornsen, "Quality assurance and control (QAC) of uncertainty models and simulations of life cycle assessment (LCA)," in *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 1995, pp. 252–257.

[3] "Single Market for Green Products - The Product Environmental Footprint Pilots - Environment - European Commission." [Online]. Available: https://ec.europa.eu/environment/eussd/smgp/ef_pilots.htm. [Accessed: 15-Mar-2021].

[4] F. J. Gómez *et al.*, "Online Life Cycle Assessment: Case Vinyl Acetate Monomer Process References 'A Systematic LCA-enhanced KPI Evaluation towards Sustainable Manufacturing in Industrial Decision-making Processes. A Case Study in Glass and Ceramic Frits Production,'" *Procedia CIRP*, vol. 48, pp. 158–163, 2016.

[5] P. Kannisto *et al.*, "Plant-wide communication architecture enabling online life cycle assessment," in *Automaatiopäivät 23*, 2019.

[6] "SULCA - Sustainability tool for Ecodesign, Footprints & LCA | SimulationStore." [Online]. Available: https://www.simulationstore.com/sulca. [Accessed: 15-Mar-2021].

[7] "Functional Mock-up Interface." [Online]. Available: https://fmi-standard.org/. [Accessed: 15-Mar-2021].

[8] "InfluxDB: Purpose-Built Open Source Time Series Database | InfluxData." [Online]. Available:

https://www.influxdata.com/. [Accessed: 15-Mar-2021].

[9] M. Overcash, E. Griffing, E. Vozzola, J. Twomey, W. Flanagan, and J. Isaacs, "Advancements in Unit Process Life Cycle Inventories (UPLCI) Tools," in *Procedia CIRP*, 2018, vol. 69, pp. 447–450.

[10] L. F. Cabeza, L. Rincón, V. Vilariño, G. Pérez, and A. Castell, "Life cycle assessment (LCA) and life cycle energy analysis (LCEA) of buildings and the building sector: A review," *Renewable and Sustainable Energy Reviews*, vol. 29. Elsevier Ltd, pp. 394–416, 01-Jan-2014.

[11] "ISO - ISO 14044:2006 - Environmental management — Life cycle assessment — Requirements and guidelines." [Online]. Available: https://www.iso.org/standard/38498.html. [Accessed: 15-Mar-2021].

[12] T. Blockwitz *et al.*, "Functional Mockup Interface 2.0: The Standard for Tool independent Exchange of Simulation Models," 2012.

[13] A. Elsheikh, M. U. Awais, E. Widl, and P. Palensky, "Modelica-enabled rapid prototyping of cyber-physical energy systems via the functional mockup interface," in *2013 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, MSCPES 2013*, 2013.

[14] K. Gulzar, R. Ruusu, S. Sierla, P. Aarnio, T. Karhela, and V. Vyatkin, "Automatic Generation of a Lifecycle Analysis Model from a First Principles Industrial Process Simulation Model," in *Proceedings - IEEE 16th International Conference on Industrial Informatics, INDIN 2018*, 2018, pp. 741–746.

[15] "Social-Impact Efforts That Create Real Value." [Online]. Available: https://hbr.org/2020/09/social-impact-efforts-that-create-real-value. [Accessed: 16-Mar-2021].

Pasi Airikka*

PI control for load disturbance rejection

Abstract: Despite diversity of available multivariable control strategies, PID controllers will remain the workhorse of industrial process control due to simplicity, availability and effective performance. Still, there are more than many untuned or badly tuned PID controllers working in real applications. And if tuned, PID controllers are most probably tuned for a smooth setpoint response. However, the primary reason for using PID control is load disturbance rejection. This paper reminds of the PID controller's importance of rejecting load disturbances over setpoint following. Load disturbance optimal PI controller tuning rules for two most typical single-input single-output transfer function models are given. In addition, the paper revisits simple and practical method for assessing the PI controller's load disturbance rejection performance in real process control applications.

Keywords: PID control, load disturbance, step response, control performance, transfer function, integrated absolute error

***Corresponding Author:** Senior lecturer, Tampere University of Applied Sciences, Tampere FINLAND, E-mail: pasi.airikka@tuni.fi

1 Introduction

Since the early 1930's, PID control has been available for process control. Its widespread usage was heavily increased when the control systems emerged in the late 1960's and 1970's due to electrotechnical development. Process control systems such as programmable logical controllers, distributed control systems and stand-alone unit PID controllers serve a platform for PID control -based applications.

Quite early, the research papers on PID controller tuning showed the importance of separating two control tasks: setpoint following and load disturbance rejection. Later, it has been shown in numerous research articles through decades that tuning a PID controller for optimal setpoint following does not correlate with optimal performance for load disturbance rejection. And this conflicting duality does not vanish even if process model uncertainties are considered by securing robustness measures such as maximum sensitivity, gain and phase or delay margins.

For closed-loop control systems, the only user-manipulated variable is typically a setpoint, that is, a reference signal. Simply by stepping the setpoint from one value to another, the closed-loop step response can be generated in real environment. As a result, the setpoint response can be visually inspected for assessing the PID controller's performance. However, this is not the only side of the coin to be looked at. The controller performance should be equally, or even more importantly, assessed for load disturbance rejection.

Most of the PID control loops operate on constant setpoints. Setpoints are seldom, if ever, changed. One of the exceptions to this rule is secondary cascade PID controllers, that is, slave controllers, which receive their setpoints from primary cascade, that is, master controllers. Also, quality-related or production rate controllers may have occasionally changing setpoints. But basically, all the other PID controllers work on constant setpoints battling against unmeasured or measured load disturbances.

In this paper, two of the most frequent and simple linear transfer function models are treated: FOPDT (First-Order Plus Dead Time) and IPDT (Integrator Plus Dead Time). The two process model types have been used to specify a simple PI controller tuning rule for optimal load disturbance rejection. Process control literature already recognizes load disturbance optimal tuning rules as given in [2], [4], [5], [6] and, therefore, the proposed method is not new although the presented tuning rules may be. However, the proposed optimal tuning rule is just an intermediate step for generating step load disturbance responses that minimize an integrated absolute control error. The responses are then analyzed in terms of e.g. maximum peak error and damping in order to give insight to what a properly tuned PI controller should visually look like.

This paper tries to bring alive clever and simple method for generating load disturbance response during normal PI control operation for visual assessment as given in [6]. The method is probably partly forgotten and partly not recognized the way it should be. The method is equally simple as setpoint stepping during closed-loop PI controller operation and, therefore, it should be equally, or more preferably, wider used than setpoint stepping.

2 FOPDT and IPDT process models

The two most typical low-order process model types that are encountered in process industry are First-Order Plus Dead-Time (FOPDT) and Integrator Plus Dead-Time (IPDT). The FOPDT process model contains three model parameters: static gain k, time constant τ and time delay θ. The model can be expressed as a Laplace transfer function:

$$P(s) = \frac{k}{\tau s+1} e^{-\theta s} \quad (1a)$$

The IPDT process model contains only two model parameters: integrator slope k and time delay θ. Also, this model can be expressed as a Laplace transfer function:

$$P(s) = \frac{k}{s} e^{-\theta s} \quad (1b)$$

The model parameters are assumed to be strictly positive, that is, larger than zero. If, for example, the static gain is negative, it should be treated as positive for controller tuning but its negative sign should be considered in parametrization of a real controller.

These model types can be used to capture the most essential behavior of many single-input single-output processes such as pressure, flow, level, consistency and even temperature processes.

The models above are typically obtained as a result of process model identification. There are several ways, some of which are rather simple, to identify the model parameters of (1a) and (1b). A nice collection of simple identification methods is given in [6].

3 Load response

The list of the existing PI and PID controller tuning rules is exhausting as pointed out in [4]. Although, optimal tuning methods is only a subset of all the methods, they are rather many. Optimality can be defined using any PID controller performance criteria and of them is an integrated absolute control error (IAE).

$$IAE = \int_0^\infty |e(t)| dt \quad (2)$$

The control error e is the error between setpoint and controlled variable such as tank level or volume flow. The setpoint is basically given by a human, or in some cases, an upper cascade controller or an advanced process controller.

Basically, the control error is due to a setpoint change or a load disturbance affecting the process. The load disturbance can be any unmeasured or measured process variable e.g. flow or level that disturbs a controlled variable. Figure 1 illustrates a simulated control error for a PI controlled closed loop when a load disturbance is changed in a stepwise manner at time of zero. The load response starts from zero and finally, settles back to zero due to an integrating controller. The colored area summed up is the IAE value. As zero control error is clearly the objective, the same criterion can be defined as a minimum IAE value.

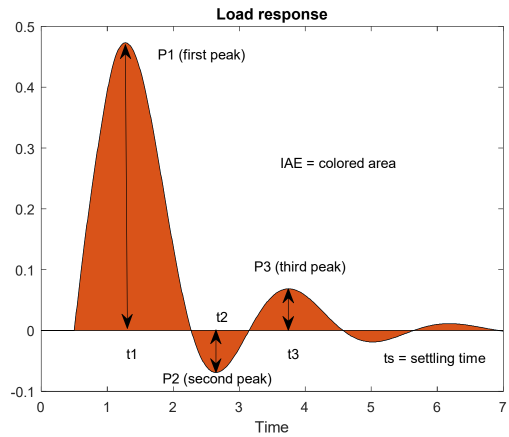


Fig. 1. Control error for a step load disturbance with IAE coloring and load response characteristics.

To characterize the step load response, there are amplitude performance indexes: first peak P_1 (maximum), second peak P_2 (minimum) and third peak P_3. Time instants for the first two peaks P_1 and P_2 are t_{max} (t_1 in fig. 1) and t_{min} (t_2 in fig. 1). The settling time is denoted by t_s. For further pattern analysis of IAE-optimal load response, the following damping ratios $\frac{P_1}{P_2}$, $\frac{P_3}{P_2}$ and product of these two $\frac{P_1}{P_2} \cdot \frac{P_3}{P_2}$ are defined. Being ratios, they do not have units. Similarly, the following time-based ratios $\frac{t_{min}}{t_{max}}$ and $\frac{t_s}{t_{max}}$ are defined.

4 Load response optimal PI controller tuning

The objective for an optimal PI controller tuning is to minimize the IAE value for any given FOPDT and IPDT process model parameterization when a unit step load disturbance acts on the process input. Quite often, load disturbance dynamics is rather similar with the process dynamics justifying its treatment as an external input to the process model for simulation. And, typically, load disturbances are slow by nature particularly allowing the usage of a step load disturbance in simulations.

For estimating optimal tuning parameters, an industrial PI controller with proportional gain k_p and integral time t_i is considered with a Laplace transfer function:

$$C(s) = k_p \frac{t_i s+1}{t_i s} \quad (3)$$

Both controller parameters are assumed to be strictly positive. In some industrial controllers, the proportional gain parameter is given in terms of a proportional band P_b parameter:

$$P_b = \frac{100\ \%}{k_p} \quad (4)$$

The objective for a load disturbance rejection tuning method is to minimize the IAE cost (2) using a PI controller (3). For optimization, two different optimization methods are used for securing reliable results. The primary method is a heuristic evolutionary random optimizer (HERO) reported in [1] which combines a deterministic gradient-like method to a population-based method. The securing method relies on Matlab function fminsearch which uses a Nelder-Mead simplex algorithm.

4.1 FOPDT process

For an FOPDT process (1a), figure 2 shows (upper left) the results of IAE-optimal tuning for proportional gain where $k_p k$ is plotted against a normalized dead time $0 < \frac{\theta}{\tau} \leq 3$. The optimal gains are plotted in black circles showing that the proportional gain changes rapidly for a normalized dead time being $\frac{\theta}{\tau} \leq 1$ whereas the gain changes significantly less for $\frac{\theta}{\tau} > 1$. This underlines a well-known fact using smaller proportional gain for processes having dead time bigger than time constant.

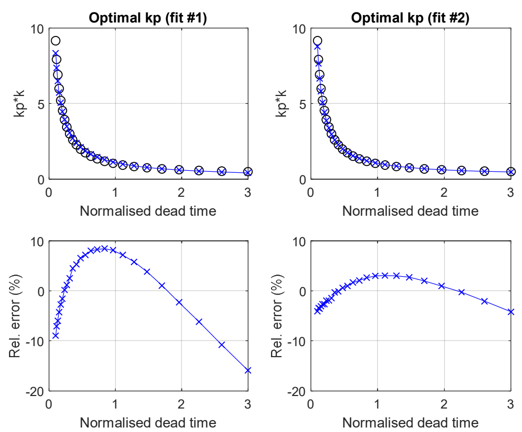


Fig. 2. IAE-optimal proportional gains $k_p k$ for normalized dead time $0 < \frac{\theta}{\tau} \leq 3$. Upper: optimal fit. Lower: relative fitting error.

Equally to figure 2, figure 3 shows (upper left) the results of IAE-optimal tuning for integral time where $\frac{t_i}{\tau}$ is plotted against a normalized dead time $0 < \frac{\theta}{\tau} \leq 3$. The optimal values are plotted in black circles showing that the integral time changes rapidly for a normalized dead time being $\frac{\theta}{\tau} \leq 1$ whereas the integral time changes linearly and slower for $\frac{\theta}{\tau} > 1$.

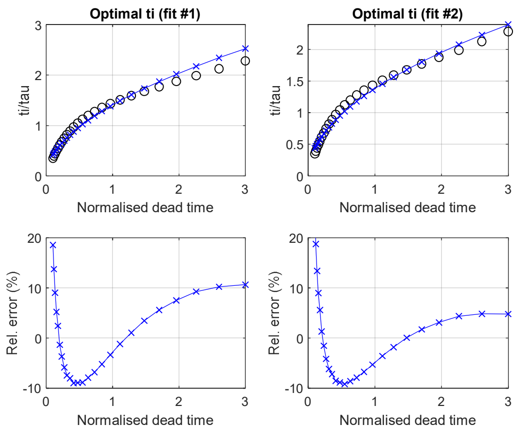


Fig. 3. IAE-optimal integral time $\frac{t_i}{\tau}$ for normalized dead time $0 < \frac{\theta}{\tau} \leq 3$. Upper: optimal fit. Lower: relative fitting error.

In both figures 2 and 3, there are two fits made for the resulted optimal controller values. The first fit (fit #1 in plots) is plotted in solid blue line with crosses in upper left plots. The relative fit error against $\frac{\theta}{\tau}$ is shown in lower left plots. The fitted tuning rules for IAE optimal PI controller tunings obtained from data plotted in figures 2 and 3 are below:

$$k_p = \frac{1.1}{k} \cdot \left(\frac{\theta}{\tau}\right)^{-0.88}$$

$$t_i = 1.41\tau \cdot \left(\frac{\theta}{\tau}\right)^{0.53} \quad (7)$$

The obtained IAE optimal load disturbance tuning rules are rather alike with those given in [2], [3], [4] and [5]. Just for a curiosity, optional tuning rules for an IAE optimal PI controller (fit #2 in plots) have a slightly better fit:

$$k_p = \frac{0.19\frac{\theta}{\tau}+0.86}{k \cdot \tau}$$

$$t_i = 1.38\tau \cdot \sqrt{\frac{\theta}{\tau}} \quad (8)$$

The fitted parameters are plotted in upper right plots in figures 2 and 3. As comparison between two fits is not easy just by looking at the upper plots, the relative fitting error is plotted in lower plots in both figures. A quick glance at the fit errors show that the rules that the given rules (8a) and (8b) provide with a better fit, especially for the proportional gain.

4.2 IPDT process

An IPDT process (1b) only contains two model parameters: slope k and dead time θ. As there is no time constant, the IAE-optimal tuning rules for an IPDT process cannot be given with respect to a normalized dead time, nor visualized as in figures 2 or 3. Instead, they are to be given straight with respect to delay θ:

$$k_p = \frac{0.92}{k \cdot \theta}$$

$$t_i = 4.1\theta \qquad (9)$$

If the IAE optimal integral time t_i was an exponential function of normalized dead time for an FOPDT process, now it is simply a linear function of dead time only.

5 Optimal load responses

In figure 1, an example of step load response was given. More step load responses are plotted for PI controlled FOPDT processes in figure 4. The plots illustrate differences in load responses due to PI controller tuning parameters being either too small or too large. The differences can be measured in maximum error, damping and settling time.

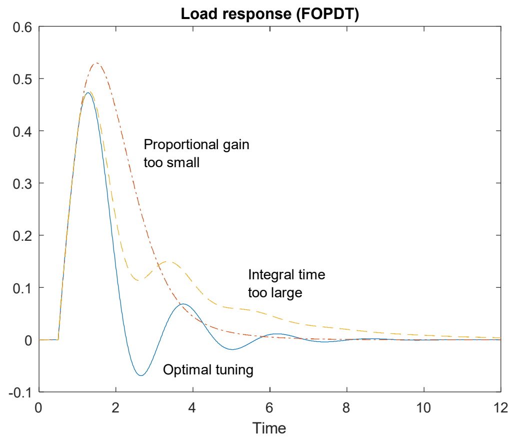


Fig. 4. Comparison of load responses for three PI controller tunings for FOPDT process.

5.1 FOPDT process

Figure 5 shows that an optimal load response has pretty much a similar shape at all values of normalized dead time $\frac{\theta}{\tau}$. This brings a question if there is performance metrics that could be used for assessing if optimal load response is achieved. Figure 6 shows that all the main metrics, IAE value, maximum load error, maximum time and setting time naturally increase with respect to increasing normalized dead time. Figure 7 shows that a ratio $\frac{P_1}{P_2}$ between the first two peaks of the load response which are maximum load error (P_1) and the minimum error (P_2) slightly increases as normalized dead time increases. A ratio $\frac{P_3}{P_2}$ between the third peak (P_3) and the second peak (P_2) decreases but the product of these ratios $\frac{P_1}{P_2} \cdot \frac{P_3}{P_2}$ stays close to 25.

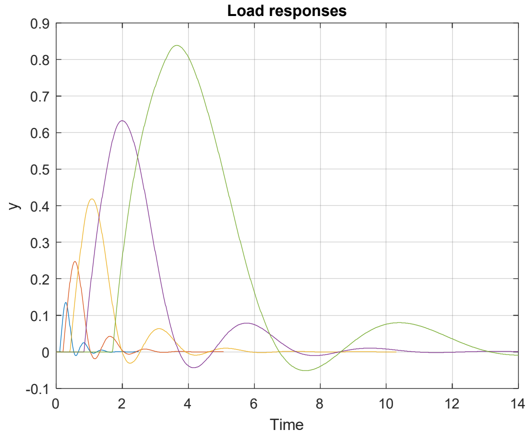


Fig. 5. IAE-optimal PI controlled step load responses for FOPDT process with $\frac{\theta}{\tau} = 0.10, 0.20, 0.41, 0.84, 1.70$.

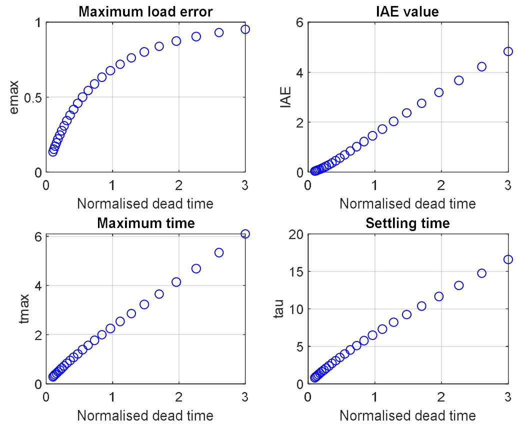


Fig. 6. IAE-optimal PI controlled step load for FOPDT process with unit static gain and time constant.

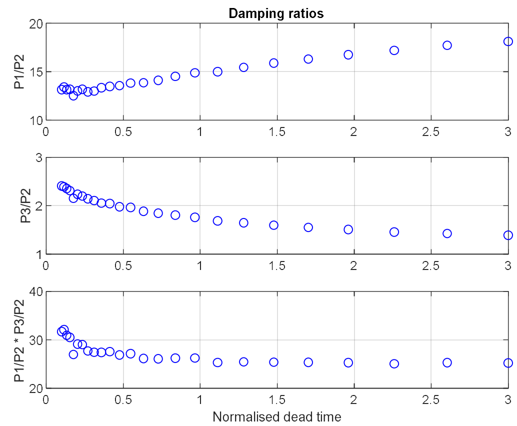


Fig. 7. Damping ratios for IAE-optimal step load responses for FOPDT process with normalized dead time $\frac{\theta}{\tau}$.

Another similar observation can be done by inspecting a ratio $\frac{t_{min}}{t_{max}}$ between maximum time t_{min} and t_{max} which are the time instants for minimum load error P_2 and maximum load error P_1. The ratio remains close to $\frac{t_{min}}{t_{max}} \approx 2.1$. One more observation is related to a ratio $\frac{t_s}{t_{max}}$ between setting time τ_{st} and maximum time t_{max}, which seems to be close to $\frac{t_s}{t_{max}} \approx 2.9$. The ratios for IAE-optimal PI controller tuning are plotted in figure 8.

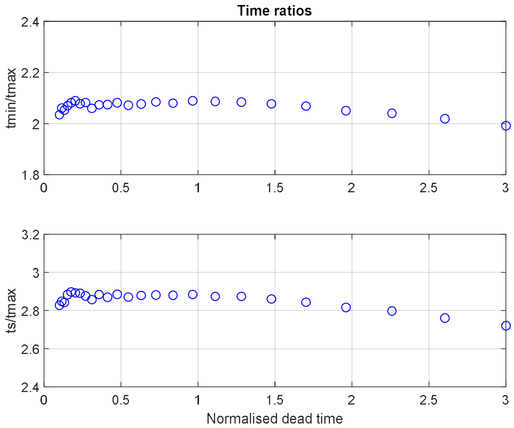


Fig. 8. Time ratios for IAE-optimal step load responses for FOPDT process with normalized dead time $\frac{\theta}{\tau}$.

5.2 IPDT process

Analysis of IAE-optimal step load response reveals similar observations for IPDT processes. The load response metrics, as given in figure 7, equally increase with increasing dead time. And, similarly, when looking at the specified ratios, they stay rather constant regardless of the IPDT dead time θ. The very same ratios for IPDT process are $\frac{P_1}{P_2} \cdot \frac{P_3}{P_2} \approx 29$, $\frac{t_s}{t_{max}} \approx 2$ and $\frac{t_s}{t_{max}} \approx 2.5$ as plotted in figures 10 and 11.

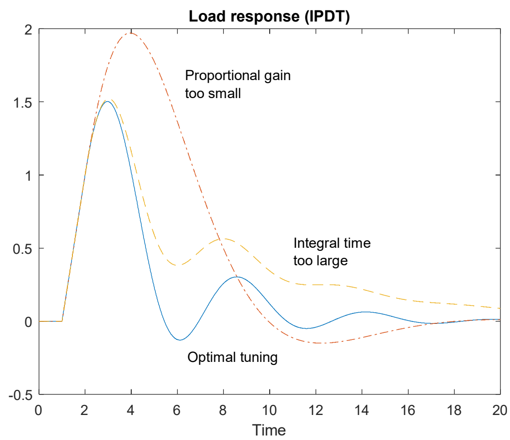


Fig. 9. Comparison of load responses for three PI controller tunings for an IPDT process.

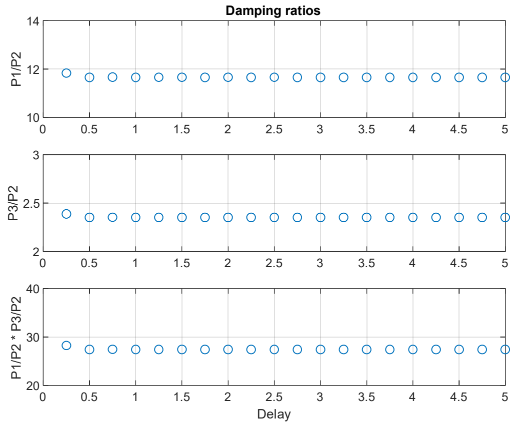


Fig. 10. Damping ratios for IAE-optimal step load responses for IPDT process with dead time θ.

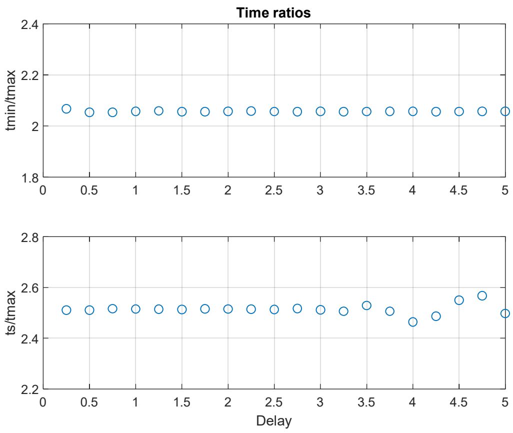


Fig. 11. Time ratios for IAE-optimal PI controlled step load responses for IPDT process with dead time θ.

6 Robustness of IAE-optimal tuning

Implementing a controller always poses a risk to lose closed-loop stability due to process model uncertainties. Therefore, robustness needs to be considered during control design. A good measure of closed-loop stability is maximum sensitivity M_s which should preferably be between 1.2 and 2.0 for a well-tuned controller. The maximum sensitivity bounds gain margin G_m and phase margin P_m which are recommended to be between 2-4 and 35-60 degrees.

Figure 12 shows robustness measures (M_s, G_m, P_m) for an IAE-optimal PI controller for an FOPDT model. The values are rather similar with those reported in [2]. The maximum sensitivity M_s = 1.9…3.0, the gain margin $G_m = 1.5 \dots 2.1$ and the phase margin $P_m = 19 \dots 30$ degrees. The robustness measures are slightly below recommendations but might be adequate if process model used for PI control design is not inaccurate and process dynamics does not change that much in real.

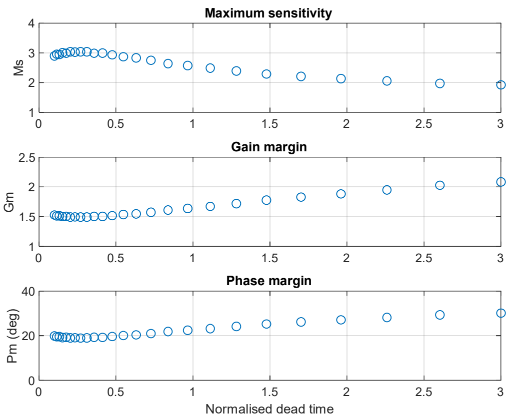


Fig. 12. Robustness measures for IAE-optimal PI controller tuning for an FOPDT process against normalized delay $\frac{\theta}{\tau}$.

7 Simple load response experiment

The observations for IAE-optimal step load response behavior call for a simple process experiment to confirm in real if the controller has been tuned for optimal load response or close to it, at least. Luckily, such an experiment exists as explained in [5]. The experiment requires a PI controller operating in automatic and in a steady state with zero error to start with. Then, the PI controller is switched off to manual mode and the controller output is manually stepped up or down. The new control value is supposed to be entered using a keyboard but, sometimes, only push buttons are available. In that case, the new value should be fed as quickly as possible to generate a stepwise change exciting the process.

As soon as the stepped control value is fed to the system, the PI controller is switched back to automatic mode. Now, the PI controller starts to compensate the synthetic load disturbance that was generated by tampering the control value. The experiment, if not disturbed by other process variables, shows a response pattern similar to those given in figure 5. Once the PI controller has eliminated the control error, the experiment is completed and the performance metrics such as maximum load error, maximum time, minimum load error and minimum time can be calculated. Based on the recorded values and the calculated ratios, IAE-optimality of the PI control loop can be assessed.

8 Conclusion

Most of the PI controllers exist for eliminating upsets due to unmeasured load disturbances while their setpoints remain unchanged for long time periods. Bearing this in mind, the PI controllers should primarily be tuned for a good load disturbance rejection instead of smooth setpoint following. Cascade slave controllers and controllers receiving their setpoints from advanced, multivariable process controllers are, of course, an exception to this rule.

Good load disturbance rejection can be achieved by computing an Integrated Absolute Error (IAE) for a step load disturbance affecting the process input and minimizing the value. This paper introduced the PI controller tuning rules for minimizing the IAE value for FOPDT and IPDT processes. More importantly, the paper reminds of a smart experiment that could be used for generating a step load response without tampering any process equipment. The experiment can be carried out basically with every PI controller just by fooling around with controller operation modes (automatic vs. manual) and entering controller output in manual.

The paper also shows how IAE-optimal load disturbance response can be recognized simply by recording and computing a few performance metrics on the resulted step load response. The values can be used for estimating optimality of the PI controller operating in automatic.

References

[1] Airikka, P. Heuristic evolutionary random optimizer for non-linear optimization including control and identification Problems, Nordic Process Control Workshop (17th NPCW), Lungby, Denmark, 2012.

[2] Ho, W. K., Hang, C. C., Zhou, J. H. Performance and gain and phase Margins of well-Known PI tuning formulas, IEEE Transactions on Control Systems Technology, Vol. 3, No. 2, June 1995.

[3] Ho, W.K, Lim, K. W., Hang, C. C., Ni, L. Y. Getting more phase margin and performance out of PID controllers, Automatica, Vol. 35, pp. 1579-1585, 1999.

[4] O'Dwyer, A. Handbook of PI and PID controller tuning rules. Imperial College Press, 3rd edition, Singapore, 2009.

[5] Shinskey, F.G. Process control systems: Application, design, adjustment. McGraw-Hill, 4th edition, New York, 1996.

[6] Åström, K.J. and Hägglund, T. PID controllers: theory, design and tuning. page numbers. Instrument Society of America, United States of America, 1995.

Pasi Airikka*

PI controller for solving simple equations

Abstract: PID control is a true workhorse for industrial process control. Its widespread use in various control applications is highly contributed by its astonishing property of removing a control error between setpoint and controlled variable. The very same feature, however, can also be utilized for other purposes where zero error is required. Solving an equation is good example of such dilemma. In this paper, the PI controller is presented as a solver e.g. for computing square roots, solving quadratic equations with real-valued solutions and linear equations of two variables.

Keywords: PID control, linear, equation, solution.

***Corresponding Author:** Senior lecturer, Tampere University of Applied Sciences, Tampere FINLAND, E-mail: pasi.airikka@tuni.fi

1 Introduction

A PID controller (Proportional-Integral-Derivative) is a backbone of industrial process control. Typically, it is implemented using a ready-made PID controller function block available on a process control system. And once commissioned, it constantly regulates a process variable as specified by removing the control error between setpoint and controlled variable if an integrating controller is used.

The PID controller has an integral controller for guaranteeing a zero error between setpoint and process output. The integral controller is based on computing an integral of the control error with respect to time. Receiving the control error as an input, it gives the time-integrated control error as an output. Combined with feedback, the simple mathematical operation of integration has an astonishing capability of driving its input inevitably to zero. The mathematical proof can be found e.g. in [6].

The amazing property of the integral controller can be used for driving any feedback signal to zero. As reported in [1], a PI controller can be used for identifying process model parameters or it might be used for matching true process measurements with simulated process model outputs as explained in [3] and [4]. In all these applications, the integrating controller plays a significant role in removing an error between two signals by feedback. One of the latest reported non-control related application of a PID controller is given in [5] where an integrating controller is used for searching prime number orders.

In this paper, the PI controller is first introduced as a calculator for computing some scientific functions such as square roots and logarithms. Then, the method of zero removal by the PI controller is extended for solving single variable quadratic equations with real roots. And, finally, the application of solving linear equations of two variables with real-valued solutions is explained. In all above cases, examples are given to highlight the way the PI controller acts as a numerical routine for solving equations or computing scientific functions.

It is worth pinpointing, that the methods given in this paper are far from being competitive alternatives for numerical routines to the existing, well-known and efficient methods for solving equations or computing functions. Numerical operations and computing time needed for computations is far too big for taking the proposed method seriously. Instead, the purpose of the paper is to show how the integral controller can be used for solving many kinds of problems, not only process control issues.

2 PI controller as a solver

A PI (Proportional-Integral) controller receives control error e as an input and returns control u as an output. The PI controller can be given as

$$u(t) = k_p(e(t) + \frac{1}{t_i}\int_0^t e(\tau)d\tau) \qquad (1)$$

with proportional gain $k_p \neq 0$ and integral time $t_i > 0$ for any time instant $t \geq 0$. The control error is a deviation between setpoint r and controlled variable y

$$e(t) = r(t) - y(t) \qquad (2)$$

In process control, the controlled variable y is typically an on-line measured physical variable such as volume flow, pressure, temperature, level or consistency. The setpoint r is the targeted value for the measurement. The controller output u is typically taken to a physical control device such as control valve or pump. Often, the control is a percentage value in the range of 0–100 %.

In this paper, the PI controller does not act on a physical control device nor receives its feedback signal from any physical sensor or transmitter. Instead, the controlled variable y is a cost function value for given controller input u. Thus, there is a true feedback closing the control loop and allowing a PI controller to act. The biggest difference to real PI controller applications is that there are no signals to be transmitted between sensors, computers and actuating devices. Instead, all the computation takes place in a computer.

Another significant difference is that there is no time-domain dynamics such as dead time or time constants involved as in true process control applications for regulating physical variables. The only relation between process input u and output y is static by nature but it can be linear or non-linear. When being the latter, it may unfortunately pose some difficulties on tuning and using PI controllers.

3 PI controller implementation

In distributed control systems, there is typically a function block for implementing a PID controller. If, however, some other platform is used, then some programming might be required to implement a PID controller for execution. In this paper, a simple discrete variant of an analogue PI controller is applied.

The PI controller (1) can discretized using a forward Euler approximation and given as

$$e_k = r_k - y_k \quad (3a)$$
$$u_k = k_p \left(e_k + \frac{1}{t_i} I_k \right) \quad (3b)$$
$$I_{k+1} = I_k + h \cdot e_k \quad (3c)$$

where a sub-index k is an integer counter ($k = 1, 2, 3, \dots$) and h is a sampling period or, in this context, a computation interval.

The PI controller can be initialized to $I_1 = 0$ but it is recommended to initialize it to a reasonable initial value for a faster convergence to a numerical solution. The reasonable initial value would be any non-zero value which would be closer to a final solution than pure zero.

In real applications, the discretized PI controller (3a-c) should be equipped with many practical features such as anti-windup but in this context those practicalities are neglected on purpose without compromising achievable results.

The PI controller always has a control direction which is either direct (positive) for $k_p > 0$ or reverse (negative) for $k_p < 0$ but cannot be both at the same time. This limitation needs to be carefully considered in solving numerical equations by a PI controller.

4 PI controller for computing simple scientific functions

The PI controller can be used for calculating many of the scientific functions such as square root, exponential functions and logarithms.

Case 1. Square root.

Compute a square root $u = \sqrt{a}$ for a given real-valued $a > 0$. Squaring both sides of the square root gives

$$u = \sqrt{a} \Leftrightarrow u^2 = a \quad (4)$$

The process (function) to be controlled by a PI controller is $y = u^2$ with controlled variable y, controller output u and setpoint $r = a$.

Example. For computing a square root $u = \sqrt{51}$, the function $y = u^2$ is plotted for $u = 0 \dots 10$ in figure 1.

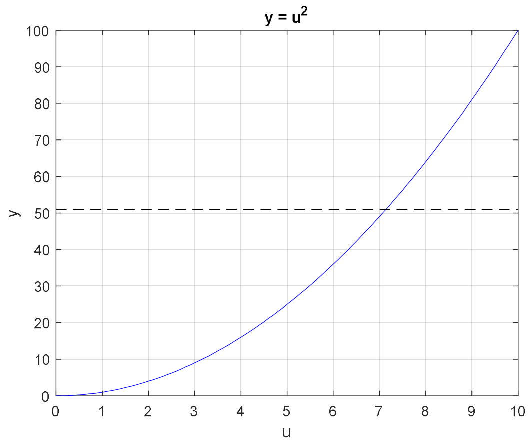


Fig. 1. Function $y = u^2$ for computing a square root $u = \sqrt{51}$.

The discrete PI controller (3a-c) with $k_p = 0.05$, $t_i = 1$, $h = 1$ is implemented for controlling the process $y = u^2$ with a setpoint $r = 51$. With $I_0 = 0$, the PI controller output converges to a solution $u \approx 7.1414$ by making the control error $e = r - y = 51 - u^2$ to zero in 21 iterations (solid in figure 2). For comparison, the integral term is initially set to $I_0 = \frac{t_i}{k_p} \cdot \frac{a}{10} = \frac{1}{0.05} \cdot \frac{51}{10} = 102$ providing with a faster convergence to a solution (dashed in figure 2).

ISBN: 13 978-952-5183-59-7

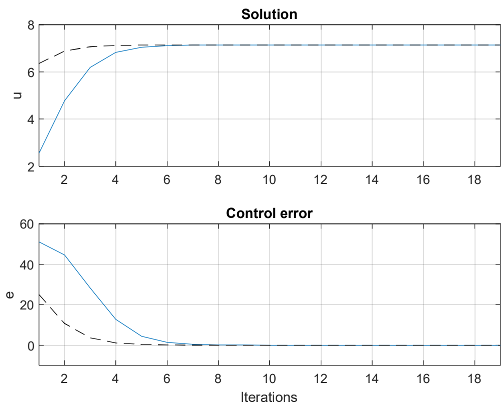


Fig. 2. PI controller responses for solving $u = \sqrt{51}$. *Upper*: PI controller output converging to a solution $u = \sqrt{51}$. *Lower*: control error $e = 51 - y$ converging to zero for $y = u^2$.

Case 2. 10-base logarithm

Compute a 10-base logarithm $u = \log_{10} a$ for a given real-valued $a > 0$. By applying exponentiation on both sides of the logarithm gives

$$u = \log_{10} a \;\Leftrightarrow\; 10^u = a \qquad (5)$$

Now, the process (function) to be controlled by a PI controller is $y = 10^u$ with controlled variable y, controller output u and setpoint $r = a$.

Example. For computing a 10-base logarithm $u = \log_{10} 51$, the function $y = 10^u$ is plotted in figure 3.

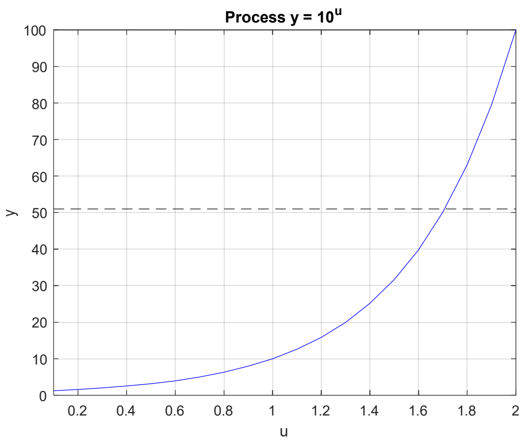


Fig. 3. Function $y = 10^u$ for computing a 10-based logarithm $u = \log_{10} 51$.

The discrete PI controller (3a-c) with $k_p = 0.01$, $t_i = 1$, $h = 1$ is implemented for controlling the process $y = 10^u$ with a setpoint $r = 51$. Without any initialization of I_k, the PI controller output converges to a solution $u \approx 1.7076$ by making the control error $e = r - y = 51 - 10^u$ to zero in 17 iterations (solid in figure 4). For comparison, the integral term I_k is initially set to $I_0 = \frac{t_i}{k_p} \cdot \frac{a}{10} = \frac{1}{0.01} \cdot \frac{51}{10} = 501$ enabling a quicker convergence to a solution (dashed in figure 4).

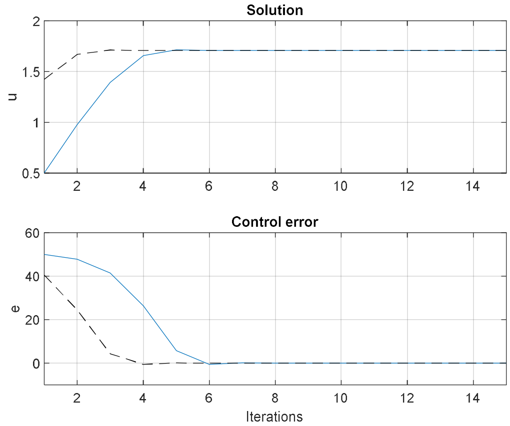


Fig. 4. PI controller responses for solving $u = \log_{10} 51$. *Upper*: PI controller output converging to a solution $u = \log_{10} 51$. *Lower*: control error $e = 51 - y$ converging to zero for $y = 10^u$.

5 PI controller for solving quadratic equations

The PI controller can be used for solving a quadratic polynomial

$$au^2 + bu + c = 0 \qquad (6)$$

with $a \neq 0$ for real-valued solutions (roots) of u only, the condition of which limits the usage of a PI controller only for cases when $b^2 > 4ac$.

The process (function) to be controlled by a PI controller is $y = au^2 + bu + c$ with controlled variable y, controller output u and setpoint $r = 0$.

Example. Solve (6) for $a = 1, b = -6, c = -16$. The condition of $b^2 = 36 > 4ac = -64$ is satisfied for existence of a real-valued solution for u. The polynomial is plotted for $u = -4 \ldots 10$ in figure 5. The roots of the polynomial are $u_1 = -2$ and $u_2 = 8$ and they are denoted by red circles whereas the turning point $-\frac{b}{2a}$ (minimum) of the polynomial is denoted by a red cross.

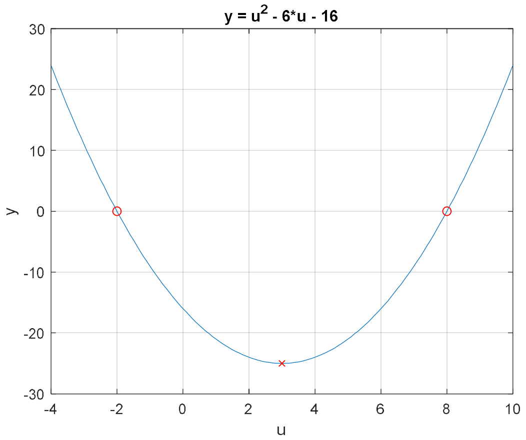


Fig. 5. Polynomial $y = u^2 - 6u - 16$ for solving $y = u^2 - 6u - 16 = 0$.

Due to the turning point of a polynomial (figure 5), the polynomial either decreases or increases for increasing u. However, the PI controller has a fixed, predetermined control direction. Therefore, two PI controllers are required, each assigned with one solution of the polynomial.

Two discrete PI controllers (3a-c) with $k_p = \pm 0.1, t_i = 1, h = 1$ are implemented for controlling the process $y = u^2 - 6u - 16$ with a setpoint $r = 0$. The other PI controller has a negative sign for the other solution. Both controllers are initialized to a turning point $u_0 = -\frac{b}{2a} = -\frac{-6}{2\cdot 1} = 3$ from where they start to operate to opposite directions due to opposite proportional gain signs.

PI controllers' outputs converge to solutions $u_1 = -2$ and $u_2 = 8$ by making both control errors $e_i = r_i - y_i = 0 - (u_i^2 - 6u_i - 16)$ for $i = 1,2$ to zero in 8 iterations (figure 6). The convergence is visualized with the polynomial function in figure 6 (right) by red circles.

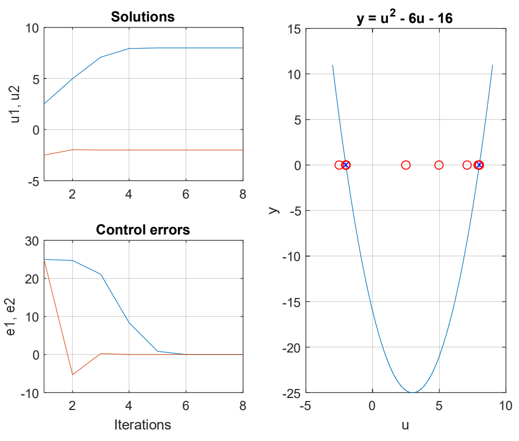


Fig. 6. PI controllers' responses for solving $u^2 - 6u - 16 = 0$. *Upper left*: PI controllers' outputs converging to solutions $u_1 = -2$ and $u_2 = 8$. *Lower left*: control errors during computation. *Right:* Polynomial with converging estimates (red circles).

The PI controllers could be used for solving polynomial of higher degrees, as well. However, the number of PI controllers required correspond the degree of polynomial as each controller is assigned for solving one root of the polynomial. In addition, turning points of the polynomial are to be computed initially making the method much too complicated for practical usage.

6 PI controller for solving a set of two linear equations

The PI controller can be used for solving a simple set of two equations with two unknown real variables u_1 and u_2

$$\begin{aligned} a_{11}u_1 + a_{12}u_2 &= b_1 \\ a_{21}u_1 + a_{22}u_2 &= b_2 \end{aligned} \qquad (7)$$

assuming the condition for the existence of the solution is satisfied. In linear algebra, the set (7) can be formulated using a matrix syntax

$$A\boldsymbol{u} = \boldsymbol{b} \qquad (8)$$

where $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. The condition for the existence of the solution can be formulated as $rank([A \mid \boldsymbol{b}]) = 2$ saying that the matrix has a full rank.

Now, there are two processes (functions) to be controlled with two controlled variables $y_i = a_{i1}u_1 + a_{i2}u_2 - b_i$, two controller outputs u_i and two setpoints $r_i = 0$ for $i = 1, 2$.

As with any multivariable control application, the input-output pairing is a concern to tackle with for designing controllers. A method often used is based on a simple RGA analysis (Relative Gain Array) as given in [2]. The same method is recommended here for selecting appropriate input-output pairs for solving the linear equations using PI controllers (7).

Example. Solve a linear set of equations

$$\begin{aligned} u_1 + 9u_2 &= 70 \\ 2u_1 + 8u_2 &= 8 \end{aligned} \qquad (9)$$

The equation set (9) has a feasible solution as the matrix $A = \begin{bmatrix} 1 & 9 \\ 2 & 8 \end{bmatrix}$ has a full rank.

The RGA method with a matrix $RGA = \begin{bmatrix} -0.8 & 1.8 \\ 1.8 & -0.8 \end{bmatrix}$ recommends pairing u_2 with y_1 and u_1 with y_2 for process outputs $y_1 = u_1 + 9u_2 - 70$ and $y_2 = 2u_1 + 8u_2 - 8$. Both processes (functions) have positive gains (9 and 2) allowing to use the PI controllers with positive (direct) control directions.

Two discrete PI controllers (3a-c) with $k_p = 0.1$, $t_i = 1$, $h = 1$ are implemented for regulating the controlled variables y_i with setpoints $r_i = 0$ for $i = 1, 2$. The controllers are initialized to zero $u_i = 0$.

The PI controllers' outputs converge to solutions $u_1 = -48.8$ and $u_2 = 13.2$ by making both control errors $e_1 = r_1 - y_1 = 0 - (2u_1 + 8u_2 - 8)$ and $e_2 = r_2 - y_2 = 0 - (u_1 + 9u_2 - 70)$ to zero in 211 iterations For visualization, only the first 50 iterations are plotted in figure 7.

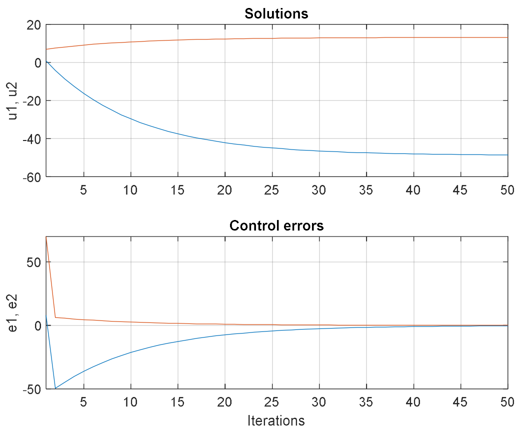


Fig. 7. PI controllers' responses for solving a set of two linear equations (9). *Upper*: PI controller outputs converging to solutions $u_1 = -48.8$ and $u_2 = 13.2$. *Lower*: control errors during computation.

7 Conclusion

The PI controller (Proportional-Integral) has an astonishing capability of eliminating a feedback error between targeted and measured variable. This amazing feature contributes to numerical integration of the control error in an integral part of the PI controller.

In process control, integral control is used for eliminating permanent errors between setpoints and measured, controlled variables. However, the method can be extended for other purposes as well. Literature shows that it has been used for model identification or searching prime numbers even. In this paper, the integral control has been used for solving simple scientific functions and equations.

A PI controller can be used as a solver for equations. However, computational effort and numerical operations required to set up a PI controller is a bit too much for practical use. In this paper, examples given were such that they are traditionally solved using algorithms and methods that do not require iterations. For example, an equation of second order has a well-known solution formula. Higher-order polynomial equations are efficiently solved using e.g. eigenvalue decomposition.

Most probably, to allow a fair comparison between traditional solvers and a PI controller, more complicated numerical problems requiring iterative approach should be introduced. Yet, increasing complexity in problems would necessarily demand more of the PI control design, the operation direction of a PI controller being just one of the issues to tackle with.

An interesting topic to study would be stability of the PI controller as a solver. During a vast number of simulations carried out for the problems presented in this paper alone, oscillations and even unstable performance was detected. Tuning parameters of a PI controller make a big difference even though there is no time-domain dynamics such as dead time and time constants involved.

An apparent reason for the stabilization issue of using a PI controller for solving equations is due to a non-linear nature of problems. Already a second-order polynomial has a non-linear element due to squaring of a variable. Nonlinearities might require nonlinear behavior or adaptation at least from a PI controller to work more reliably and without compromising stability.

References

[1] Airikka, P. Continuous-time parameter identification using PI controllers. Nordic Process Control Workshop, Oulu, Finland, 2013.

[2] Bristol, E. H. On a new measure of interaction for multivariable process control. IEEE Transactions on Automatic Control, 1: 133–134, 1966.

[3] Friman, M., Airikka, P. Tracking simulation based on PI controllers and autotuning, 2nd IFAC conference on advanced PID control, Brescia, Italy, 2012.

[4] Friman, M., Airikka, P. Tracking simulation method, Patent US20130116802A1, 2010.

[5] Klán, P., Optimized integral controller for searching

prime number orders.

[6] Åström, K.J. and Hägglund, T. PID controllers: theory, design and tuning. page numbers. Instrument Society of America, United States of America, 1995.

Olli Ventä*

Digitalization of Industry – past, present and future

Abstract: This paper makes an overview of digital transformation in heavy industries. A short look back to the development of digital industries is given but the major emphasis is given to the trends, technologies and visions of modern digital technologies relevant to industries. The presentation also brings out the major takeaways from several European research agendas, strategy papers, and white papers – where the author has been either as a leading contributor, a member of a team, or giving comments.

Keywords: digitalization of industries, European industrial strategies, digital transformation

***Corresponding Author:** Olli Ventä,
Finnish Automation Foundation,
E-mail: olli.venta@outlook.com

1 Introduction – Evolution of Industries

When the German Industrie4.0 was introduced in 2011, they also highlighted the progress of modern industrialization by **revolutions of technologies**. At first, mechanization in 1700'ies created the first generation, a transfer from human or animal driven production to machines. In around 1870, electrification brought industries into the second generation. Thirdly, introduction of computers, on the 1970'ies, meant the third generation which saw **digital automation systems** at the heart of production. And thereafter since 1970, every decade has envisioned automation as a triumph of **digitalization**.

Gradually, the architectures of digital automation systems took a well stabilized form like in Fig. 1. On bottom and at level 0, there is the actual physical or chemical production process. It has been commonplace to categorize production processes into continuous, discrete, or batch processes. Above level 0 are sensor and actuator devices comprising level 1. Nowadays, these devices may be quite intelligent, often connected to internet. Above sensors and actuators and at level 2, we see unit process controllers, typically and traditionally implemented by the so-called PLC (programmable logic controller) computers designed specifically for real-time and critical production process manipulation.

At level 3, since 1980'ies, we have seen the so-called DCS (Distributed / Digital Control System) computers managing the manufacturing extending to production lines. Along the years, DCS systems were accompanied by other important functions of manufacturing operations. Level 4 started to consist of the diverse ERP (Enterprise Resource Planning) systems.

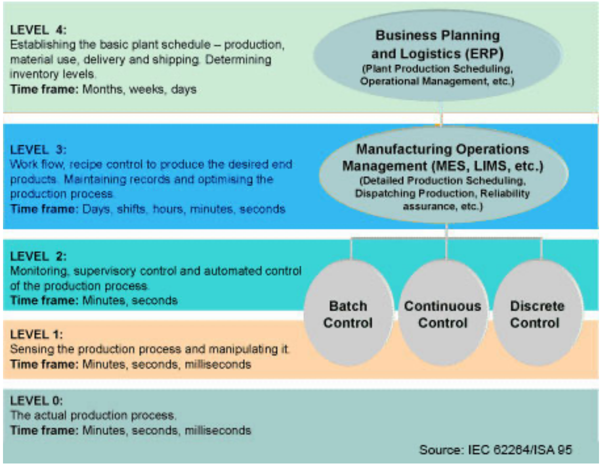


Fig. 1. Function levels of modern digital industry [https://isa-95.com].

Levels 1-3 were for years rather separate and isolated **factory automation system**, whereas, the level 4 systems were regarded as many kinds of **factory information systems**.

The decades from 1970, introduced on the scene more and more functions that perhaps existed earlier as manual functions but were gradually digitalized. The **extent of automation expanded**, and at the same time, the originally isolated level 1-3 systems found more and more useful connections creating in the end massive and complex enterprise-wide information systems.

In around 2011, the famous **Industrie4.0** was published implementing or envisioning the wider scope the manufacturing, listing the preferred or accepted subsystems and their mutual interactions or standards, and thus defining a single gigantic digital framework for all kinds of manufacturing systems. Industrie4.0 was introduced as a national initiative for German discrete manufacturing industry, especially for car manufacturing. The strategy involves the strong customization of products under the conditions of highly flexible mass-production. Throughout following years, nearly all European countries started to guide their respective national initiatives according to Industrie4.0. The latter part of 2010'ies, saw also extensions of Industrie4.0 to continuous processes. The degree and extent of digitalization has been so massive and significant that the era of 2011 and beyond is called the fourth revolution on industries.

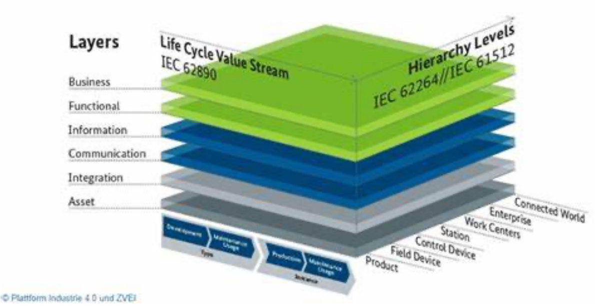


Fig. 2. Dimensions of Industrie4.0.

As we all know, industrialization arrived to Finland rather late but, for the recent 150 years, the speed has been superior. Also, automation caught up effectively, and a very good description of those times can be accessed in the three *Teollisuusautomaation vuosikymmeniltä 1-3* (Automation History) books created by *SAS* (Finnish Automation Society). The books cover the years up to about 2003. [order in: https://www.automaatioseura.fi/julkaisut-kirjakauppa/kirjakauppa/]

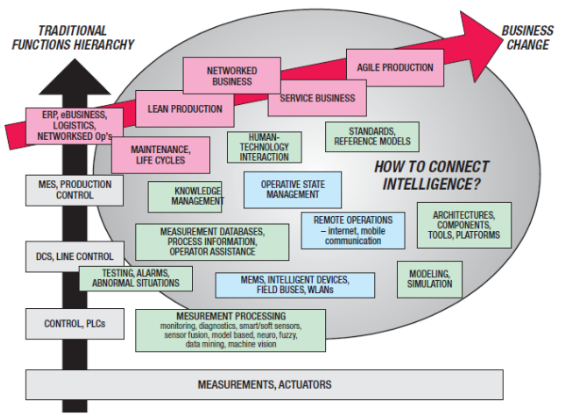


Fig. 3. A crystallization of the extensive strategy building effort of the Intelligent Automation Systems technology program (*ÄLY, 2001-4*). [2]

Entering the first decade of the second millennium brought us the national Intelligent Automation Systems technology program (2001-4) [2]. One of the most memorized activity during the program was its extensive strategy building work throughout the program. The adjacent Fig. 3. summarizes the past, present, and future of automation, in about 2003-4. The expansion of automation was clearly happening and emerging. The corresponding topics were also largely present in the projects of the program.

Along with 200X'ies, Finland got one of its most significant innovation instruments, the *SHOK*s (Strategisen huippuosaamisen keskittymä, Strategic Centres for Science, Technology and Innovation, https://fi.wikipedia.org/wiki/SHOK). *FIMECC* (Finnish Metals and Engineering Competence Cluster Ltd., https://fi.wikipedia.org/wiki/FIMECC) was one of the most successful of these new initiatives, and promoting service business had become a strategic business innovation for a growing number of Finnish companies. In 2011, we envisioned the emerging **remote service technologies** by the adjacent picture. The trends visible already in 2003-4 were seen in this picture, as well, even strengthening.

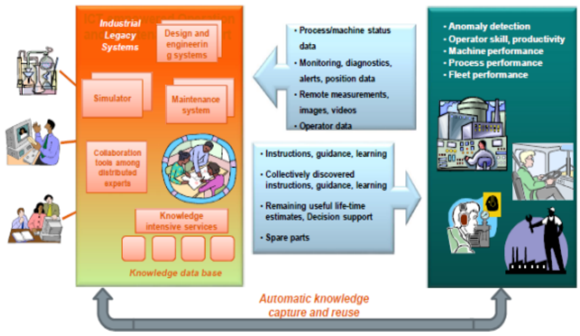


Fig. 4. Visioning industrial service business at FIMECC, in 2011.

In 2011 in Germany, the first drafts of the emerging concept of Industrie4.0 were edited. Fig. 2 in above, actually represents **RAMI**, The Reference Architectural Model for Industrie4.0 [3], indicating the dimensions of modern factory system (hierarchy levels, life cycles, and physical extents starting from product or device and enlarging to enterprises or even the connected world). In overall, RAMI defines the business processes, functions, data, communication, integration, and interactions with the physical world.

Topic wise, the contents of Industrie4.0 can be summarized as in Fig. 5. The inner circle of topics indicates the central properties of Industrie4.0 based systems, e.g., service orientation and product personalization. The outer circle lists the most significant enabling technologies that are needed to implement these properties, e.g., simulation and modeling, cloud computing, and cybersecurity. There are many presentations of topics and technologies for the digitalization of industries; here they are all in one picture.

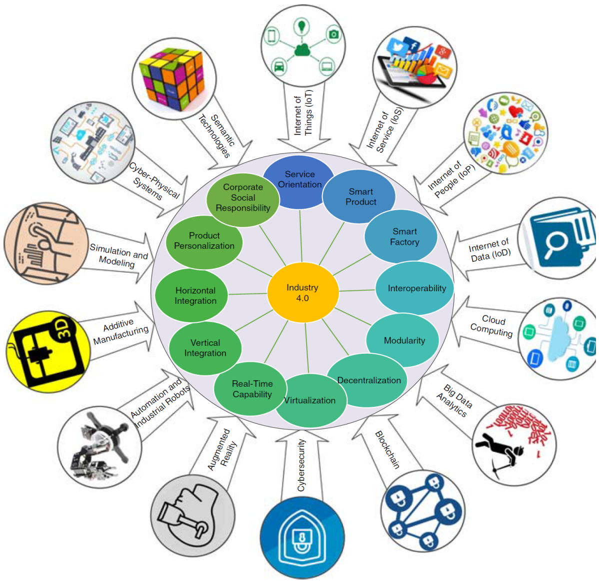


Fig. 5. Topics and enabling technologies of Industrie4.0.

After 10 years of intensive engineering and development, the industrialization domain is regarding to be on move to the fifth level of revolutions, **Industry5.0**. (see, e.g., [4]). Fig. 6 below, outline the transfer of topics or emphases. One might argue that Industry5.0 represents rather a continuum from Industrie4.0, just highlighting sustainability, human-system interaction, and the so-called resiliency.

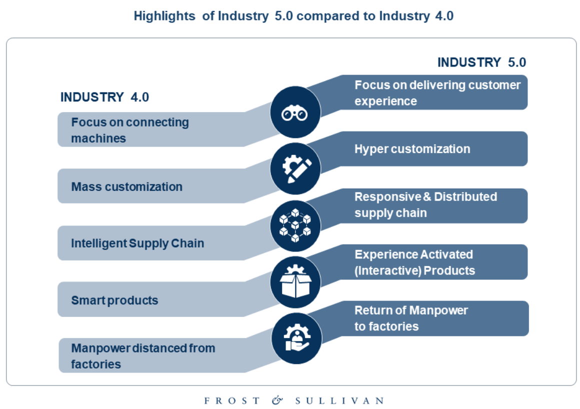


Fig. 6. Digitalization of industries on the 2020'ies – the fifth generation of industrialization.

2 Trends

2.1 Establish trend: Supply chains, networked business, globalization

In the past, at least large companies were proud of having everything they needed for business within their own factory or hands. Not so anymore. When talking about **networking**, we usually put a manufacturer or product in the middle, and place subcontractors, part suppliers, service providers, etc., to the left or upstream in the **material or value chain**. Whereas, to the right we place chains of stakeholders building up the network paths from the manufacturer to the market or end-customers. There are many forms of networking concepts that have reached a doctrine like maturity. When giving responsibility of a business function out of the company, we talk about **outsourcing**. When moving a factory out of the country, we talk about **globalization**. When supporting a customer company in maintenance. spare parts, or design, we talk about **industrial service business**. Often enterprises have geographically or topically distributed facilities, internally, and it is practical to regard it as a networked business, as well. Digitalization has played a critical role when companies have been developing their networked businesses.

It is not always evident who is strategically the most significant player in the network. Often it is the manufacturer or product assembly company (e.g., car manufacturer) but some actors immediately before the product assembly may govern more about digitalization in the chain. Examples include engine provider and factory design main contractor. Car industry calls these **first-tier** companies, and consequently the next level companies the **second-tier**, etc.

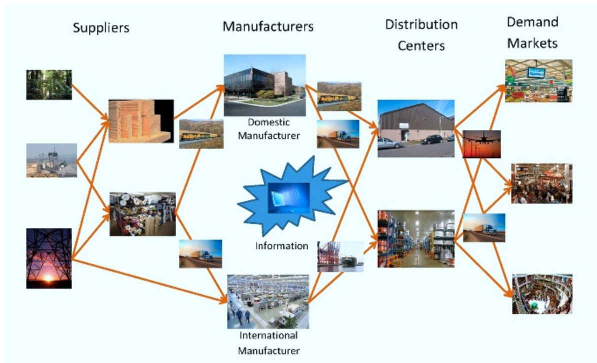


Fig. 7. Networked business. (Source: Wikipedia)

Recently, companies of developed countries have been **offshoring** manufacturing to cheaper countries. The primary motivation was in the first-place cheaper production and some other costs or benefits, sometimes availability of work force. This advantage has been vanishing, or offshore destinations have been moving from China to India, Vietnam, or Africa. Looking from Finland, a similar situation has been in regard to former Eastern-European countries. A more solid argument is **closer to market**, since typically the large populations have become important customers. The progress in automation and digitalization tend to make other stages than manufacturing in the value chain more valuable, so that actually it does not matter anymore where manufacturing is carried out. Design, innovation, product properties and ownership on one end, and remote services, customer intimacy on the other end matter more. The skilled work force in old

ISBN: 13 978-952-5183-59-7

Europe or America has remained a strength and so has emphasized high-end, one-of-a-kind production.

Covid-19 has brought at least disturbances to supply chains. Redundancies in supply chains has always been an advantage, and uncertainties have caused some call-backs of offshored factories.

2.2 Established trend: Life cycles

From product, system, or service point-of-view, conceptualizing the evolution in time, i.e., from an early idea through design and production to marketing, sales and end-use – is called a product, system, or service **life cycle**. Moving along a life cycle trajectory is not linear alone. E.g., feedbacks from manufacturing issues to product design are commonplace, as well as, **feedbacks** from subsequent stages to actually all preceding stages common. Customer experience or customer feedback deserve special attention here. Managing unexpected feedbacks may easily represent 50% of the engineering effort. A more recent and ever more important feedback type is **recycling**, closing material loops, or product item **reuse**. A note in terminology is due here, meaning industrial service business remotely downstream in networks, is often called **life cycle business**.

In the past and often today, stakeholders in various nodes in the network or software systems in different life cycle stages, have their own engineering, information or operations systems – from a multitude of system vendors or software houses. Although standardization has been intensive and initiatives like Industry4.0 has many rectifying or unifying elements, **interoperability gaps** have painfully remained a serious obstacle for efficient and fluent co-operation, or major source of useless, error-prone and expensive rework.

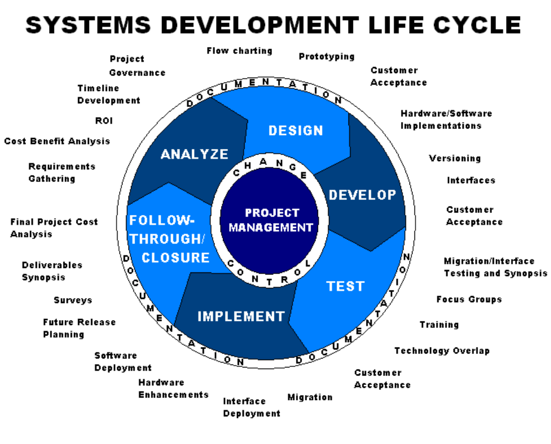


Fig. 8. Life cycle. (Source: Wikipedia)

2.3 Established trend: One-of-a-kind

Industrialization was once a single product business, one size fits all. But already for some decades, companies have been looking for advantages by one-of-a-kind production, mass-customization, product families, etc. This creates a lot of complexity to the entire production network and through life cycle, and to all their digital systems. The complexities are repeated in multiple designs, multi-path production lines, exponentially multiplied amounts of testing, customer follow-up activities, etc. One-of-a-kind production has also been a major characteristic in Finnish industry. Again, effective digitalization is a means to manage these challenges and complexities.

Fig. 9. One-of-a-kind vehicles. (Source: Wikipedia)

2.4 Established trend: Function hierarchies

This may be called a doctrine of 50 years, or as long as we have had computers in production. It has proven to be a very useful outcome of an efficient **divide-and-conque**r engineering. In the past, systems mentioned in Fig. 10 were almost separate information systems but nowadays they are very connected and extending where appropriate or necessary across company borders, production lines, etc. Industrie4.0 is built on this thinking.

Today's smart factories often keep these functions, or introduce new ones, but prototype more flexible architectural structures than the older rigid hierarchy. This is especially true for new computing platforms of advanced edge, fog, cloud, distributed, fast communication (5G) technologies.

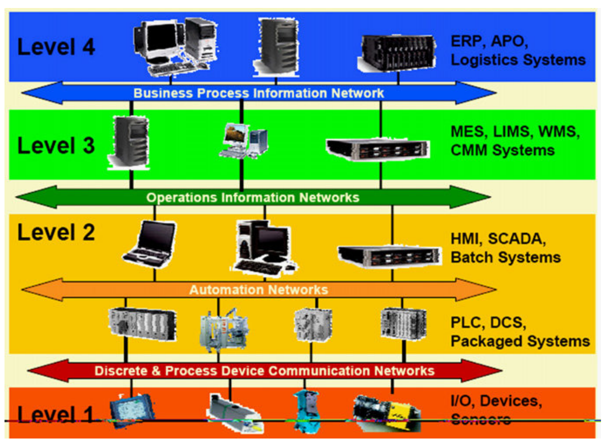


Fig. 10. Established factory automation and information

system. (Source: Wikipedia)

2.5 Recent trend: Sustainability, saving the planet

Nearly 200 countries have committed to the Paris Agreement on climate change to limit global warming to below 2ºC. Rapid transformation out of greenhouse gases (CO2, etc.) or fossil use in all sectors is required. Many countries have set even more ambitious targets. Climate awareness is so widespread that industries do not want to spend any more time in the old ways of doing. If technology allows, shifts to carbon-free production and products will occur faster than anticipated, although at the same time we must admit that hard research and innovations are desperately needed at certain points. Challenges are global, so are solutions, including also today’s poorest economies.

Digitalization could have a great bearing, or is actually absolutely necessary, in reducing environmental impact through sustainable manufacturing, energy and resource efficiency and applying circular economy strategies (eco-design, repair, reuse, refurbishment, remanufacture, recycle, waste prevention, and waste recycling, etc.).

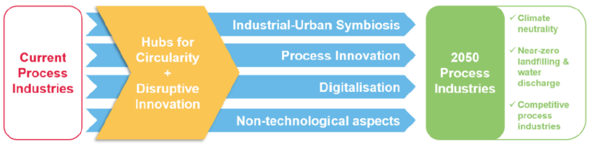


Fig. 11. Combining of means to reach the conditions of saving our planet. [7]

Environmental effects cannot be achieved by a single means and by optimizing discharges pointwise. Global value chains need to be considered thoroughly. Probably progressing towards zero emission conditions do not occur simultaneously, so net-effect analyses and controls may offer temporary significant flexibilities.

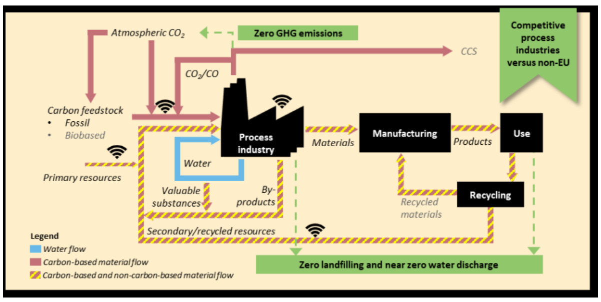


Fig. **12**. Carbon neutrality, zero waste, etc., are achieved by improving technologies and considering the entire value chains or loops. [7]

Concerns about environmental threats have been alerted for years, and industries to a growing extend have been reacting to the challenges for years. For instance, the Finnish pulp & paper industries have made a tremendous progress in reducing its environmental burdens since 1960’ies. Still, it is actually surprising how quickly the attitudes and scenery have been changing. Energy and process industries by large are now moving out of fossil fuels. Burning peat or even wood is becoming a suspect in Finland. Transfer to hybrid and electric cars is happening quicker than anticipated few years ago. Industries in the end have the resources and know-how to save the planet, and many encouraging things are happening, faster than expected.

3 Discussing selected enabling technologies

3.1 Artificial Intelligence

Artificial Intelligence (AI) as it is understood today is, on one hand, the most celebrated technology family we have, but on the other hand, it is a very good example of a promising concept of decades’ lifetime, steadily growing but annoyingly postponing the break-throughs. Early concepts and experiments were carried out already in the 1950’ies. In 1970’ies, AI became widespread in research but it primarily meant expert systems, rule-based system, LISP programming, etc. 1980’ies saw the rise of Artificial Neural Networks, leaning on effective data-driven machine learning. Already at that time, **Finland was regarded as a superpower of neural networks**. The research started in Helsinki Univ. Technology (Self-Organizing Maps) and expanded fast to all universities. Signal processing, image processing and pattern recognition (now categorized as parts of AI) had already become popular themes in Finnish research. Industrial experimentation started in 1990’ies, and has been growing since. In 2010’ies, the data-driven technologies became the main stream, and the technology family got the name Artificial Intelligence from the past. Since amounts of data increased (time series, image, video, free text, etc.), the term **‘big data’** was adopted.

Condition (or quality) monitoring and AI control have been the first and most widespread applications in industry. Since 1990’ies, neural or AI control took a major share of the automation research in Finnish universities. In condition monitoring, AI methods offered a means to provide maintenance services to industrial customers based on remote measurements, remote analytics, remote predictions, etc.

AI will impact several main areas:

- By weaving AI into the design, manufacturing, pro-duction and deployment processes, productivity can be raised.
- By using AI to increase autonomy, higher operat-ional flexibility can be achieved.

- By using AI to improve usability of products and services (e.g. by allowing greater variations in the human-machine interaction), the user value can be increased.
- By using AI for supporting complex decision-making processes in dynamic environments, people can get help in situations of rising complexity (e.g. technical complexity, increasing volatility in markets).

Computation platforms for AI are in dramatic change. The near-process intelligence is being built both on smart IoT cards or more generically, on local edge computing hardware interacting directly to the physical process and to the internet, or recently to or as part of 5G. What is not meaningful to compute locally will be computed centrally in the cloud infra. An extreme application for the future will be supporting autonomous driving locally in the edge computer(s) embedded in cars, 5G node computers in the communication network base stations, and the rest in the cloud. Similar experiments are being prototyped in factories. – AI algorithms are easily available in software libraries. Special hardware is being developed for ever faster AI computations, in the future perhaps in quantum computers.

In spite of remarkable progress, advances are often slower than expected. Few reasons can be pointed out:

- AI algorithms remain a **black box**. It is often difficult to understand and manage their behavior and performance, especially when things change or new data arrive. Their so-called **transparency** is weak. The algorithms are often tuned to maximum but to achieve reliably a predetermined performance is not distinctive to AI.
- AI modules are often **isolated** from their wider industrial context. The industrial standards define seldom statistical entities or uncertainties. AI experts often have little knowledge and experience on industrial processes, manufacturing, businesses, or mindsets.
- AI algorithms should be more commonly combined with other methods or **engineering**, as a genuine part of them, not as an isolated subsystem to be easily left aside.

3.2 Digital twins, mixed or augmented reality, telepresence

A digital twin is a dynamic digital representation of an industrial asset enabling companies to better understand and predict the behavior and performance of processes and. Nowadays, connectivity to cloud allows us an unprecedented resource. Simulation capability is currently a key element to European process, manufacturing, and machine industry to increase competitiveness. In Industrie4.0, modelling and simulation play a key role. A holistic engineering approach is required to span the different technical disciplines and prove the end-to-end engineering across the entire value chain.

Telepresence technologies can be considered as the predecessor for Extended Reality (XR) presence. The combination of XR and 5G offers a great innovation potential. The main driver is the enhancing of competitiveness through better productivity, improved worker safety, and better quality. The industrial applications have followed the prospects created by the gaming industry and consumer applications.

Besides **virtual commissioning**, modelling and simulation responds to a wider extent to many kinds of digitalization challenges:

- **Understanding, explaining, and visualisation of physical or real-world** phenomena of products, production, businesses, markets, etc.
- Helping designers to perform their core tasks, i.e., **studying alternative designs**, optimising solutions, ascertaining safety, providing a test-bench for automation and IoT solutions.
- The effects of **changes** can be safely and more extensively experimented in advance in a virtual domain than by using real plants, equipment or even mock-ups.
- Simulators offer versatile environments for user or operator **training**.
- Simulators may be used online and parallel with its real counterpart to **predict** future behaviour and performance, to give early warnings, to outline alternative scenarios for decision-making, etc. In spite of years of research, such **tracking simulators** are still in their infancy, at least in industrial use.

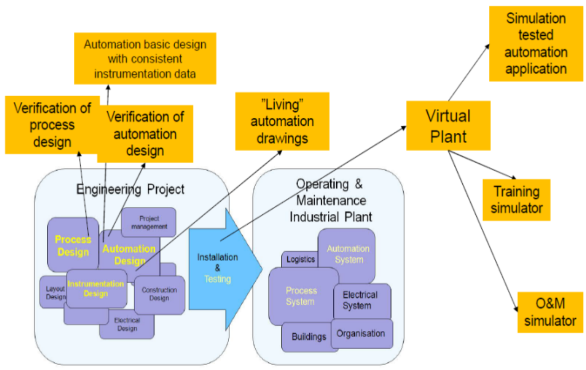


Fig. 13. Virtual commissioning in process industry.

The idea of modeling and simulation is not new, but it is not complete yet, or reaching out industries is in the half way. Some observations for future research or development are due:

- **Virtual commissioning**: It often takes substantial amount of time and effort to build a comprehensive dynamic model for the entire plant, production line, machine, or the simulator modules for the whole. For most purposes, today's modeling and simulation techniques can be accurate enough, and the special hardware for obvious increased computation complexity is available, if ordinary PC is not fast enough. With good understanding, simulation granularity can be adjusted, and partial simulations are also useful.
- **Tracking mode simulation**: Possible today but needs good models, professional engineering and usually extra computing power.
- **Generating simulators** from other design documentation or measurement data (laser scanning, etc.). Speeds up simulator design, and the technology has recently become available.
- Digital twins **combined with data-driven** (AI, etc.) models. Development and piloting needed, but holds promises for increased performances.
- **Interoperability** remains a challenge here, too. Modelling and simulation remains easily an isolated exercise in a company. In the context of design tools (CAD, etc.) interoperability exists. When developing simulators, interoperability must be kept in mind, and conversely, when designing processes and machines, engineers must bear in mind that the target they are working on, will sooner or later be simulated, too.

3.2 Design and Architecture

Digitalization has become ever more vital. It has also enabled the creation and exploitation of the digital continuum of engineering and operative computer-based systems. For instance, requirements analysis and conceptual design are carried out with the aid of specific computer-based tools. Product and system design are carried out by a multitude of software tools, often referred to collectively as **CAD** (computer-aided design) tools, with each application field (oil and gas, energy, pulp & paper, discrete, etc.) using its own variations of tools.

Engineering software has also steadily evolved into extensive product life cycle management (**PLM**) systems with CAD and product data management (**PDM**) subsystems at their core. Engineering is, in lifecycles, followed by manufacturing or construction and delivery, managed again by specific engineering or manufacturing control software. Thereafter, the product or system enters its intended industrial use and, depending on the case, effective operation is governed, for example, by sensor or actuator systems, machine or process control systems, wider automation, condition monitoring, or quality management systems.

At a higher level, production planning, enterprise resource management (**ERP**) systems, and even cross-machine management systems and networked business management systems, may be used. This continuum of engineering and operative computer-based systems provides the foundation for the digitalization of industry. Nevertheless, it is more acute than ever to really talk dually about virtual plants or machines and real plants or machines. The volume, value and importance of the digital, virtual realm is increasing dramatically compared to physical plants and machines.

Two overall assets are essential in design and engineering: (1) effective architectures and platforms at all levels of the design hierarchy and (2) structured and well-adapted design methods and development approaches supported by efficient engineering tools, design libraries and frameworks. Future systems will be intelligent (using methods from Artificial Intelligence or else), highly automated and even autonomous and evolvable, meaning their implementation and behavior will change over their lifetime. Such systems will be connected to, and communicate with, each other and the cloud, often as part of an integration platform or a system-of-system.

Design and engineering projects are challenged by **pressures of time** (faster throughput), increased complexity (extent, volume, interactions, multi-technology), and **design quality**.

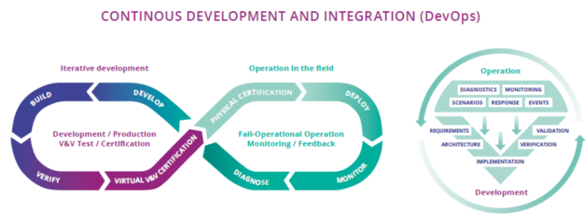


Fig. 14. Continuous development processes. [6]

Finnish industrialization started in process industries, most notably forest, mineral and energy industry. Along with decades, significant chemical and petrochemical plants became on scene, accompanied by various machine building companies, from very big to very small. Export of machines has been prevailing, and also designing and building factories abroad, i.e., global plant engineering.

This strong engineering background has brought our companies to high international level. On top of that, and based on the above-mentioned recent progresses, we can highlight the following broad focuses of development:

- **Virtual engineering.** Design processes must be expanded to enable virtual engineering on all hierarchy. Central to this approach are “digital twins”, which capture all necessary behavioral, logical and physical properties of the system under design in a way that can be analyzed. This allows for optimization and automatic synthesis. Supporting methods include techniques to visualize V&V and test efforts, as well as, sensitivity analysis and robustness test methods for different parameters and configurations. Test management within such virtual engineering processes must be extended to also cover all layers of the design hierarchy, and be able to combine virtual and physical testing. To substantially reduce design effort and costs, a second set of supporting methods deal with the automatic generation of design artefacts such as design models, automatic scenario, use-case and test vector generation, generative design techniques, design space exploration, etc.
- **System and component design (methods and tools).** To fully enable virtual engineering, design processes have to switch completely to model-based processes where models may be constructed using data-driven methods.
- **Lifecycle-aware holistic design flows**. Closing the loop, i.e., collecting relevant data in the operation phase, analyzing it and feeding it back into the development phase.

Professional design is not possible without a legion of digital tools and the underlying framework that enables smooth switching, interaction and communication. Interoperability and standards are essential. A lot of expectations have been laid on European digital single market which is difficult to achieve. Every now and then, we can read of surveys explaining how much extra work is wasted for incompabilities, to justify the standardization initiatives. Actually, errors caused may be even more severe. On the other hand, engineering industries are used to deal with the issue: they choose compatible tools and applications or they have tested and proven interaction softwares. Often, it is the knowledgeable paying customer who has strong opinions about standards, etc., and then the service providers must adapt. Managing with standards is also an obstacle for newcomers, start-ups, etc., to join big projects, or vice versa, a questionable business tactics for those who are so-to-say in, or have market shares to protect.

Digital twins are commonly characterised by modelling and simulation (the finite element method, FEM, computational fluid dynamics, CFD, etc) or virtual or mixed reality techniques, and their numerous applications. However, the product processes, manufacturing design and management of the operative lifetime of a product or factory is much broader. Typical examples of these are: managing the multi-technologies (mechanical, electronics, electrical, software); safety, security and reliability engineering; managing interactions with the contexts of the target (humans, environment); managing testing and quality; the various types of discharges or footprints; managing projects, logistics, supply chains, etc. These tasks are increasingly being managed by software tools and systems, and through the use of standards, regulations and engineering handbooks, which generally require extensive domain knowledge and experience.

The respective engineering disciplines are well distinguished, developed and understood. Key examples here – such as factory design, electronics design, engine design and car design – are well known and significant as regards success. These disciplines are going through a tremendous and demanding digitalisation process, and are sometimes called the “other twins” to underline their importance and high value. A narrow focus on digital twins will certainly play a growing role in implementing the concomitant increase in types of “other twin”.

There is also a notable discipline called “systems engineering” [19], which describes both aspects and the whole of the instantiated subfields such as factory design and engine design. Similarly, many notable software tools – such as product lifestyle management (PLM), supply chain management (SCM) and CAD – are actually families of tools with significant versions for the actual subdomains:

- parallel joint engineering of products, processes, safety, security, cybersecurity, human factors, sustainability, circular factors, etc.
- mastering the deep linkage and complexities in multiple engineering domains and technologies, along with product and process lifecycles in the digital domain.
- multiplying the engineering extent, efficiency and quality in the digital world.

3.3 Software Engineering

Software clearly represents an enabling technology for process, manufacturing, and machine industries. Software has definitely become a core, and the staff of engineering offices or departments of companies are mostly software or automation engineers. A company may have made a strategic decision not to get involved in software engineering but instead to outsource everything or at least to buy software as tools, modules, libraries, or as services. Or conversely, software has become a growing asset of company’s technology base.

Even though, we have had computers for 50-80 years, modern time has still decided to emphasize **digital transformation** [16]. Even though there are outstanding position papers, strategies and visions about today's digitalization, in the end there is no common definition for digital transformation. But it certainly means architectures, processes and technologies. Digitalizing industries is the many things delineated in this presentation, and beyond. Digital transformation is not something that you get. It is a thing that **every company has to define uniquely for itself**.

The intention here is not to define a complete doctrine or profession for software engineering, directed to industrial needs. Besides the sole software engineering itself, it should include knowledge and experience on industrial engineering of the sector of the company, on the businesses, on customers, etc. This is clearly much more than any university degree can accommodate, so it is a lifelong learning, and companies must understand and cherish it. As for individuals, it is also more than true that becoming competitive as a company needs years of attention and evolution in software engineering, and it is never ready. One practical source of basics and advanced self-studies are so-called **Software Engineering Handbooks**, e.g., [17], which give extensive and up-to-date introduction and cross-section for further reading.

Modern software used in products such as cars, airplanes, operation robots, banks, healthcare systems and the public sector comprises millions of lines of code (growing). To produce this level of software, many challenges have to be overcome. How do we develop this software and simultaneously manage its complexity? How do we ensure the correctness and security of this software, as human wellbeing, economic prosperity and the environment depend on it? How can we guarantee that software is maintainable and usable for decades to come? Finally, how can we construct the software efficiently, effectively and sustainably? Despite the fact that such software impacts everyone everywhere, the effort required to make it reliable, maintainable and usable for longer periods is routinely underestimated.

I choice of software engineering topics that would need R&D attention, is extracted from the recent SRIA of ECSEL. [6]

From software engineering to systems engineering. Developing professional software is multi-disciplinary. There is a whole ecosystem of models (e.g., physical, mechanical, structural, software and behavioral) describing various aspects of a system.

Integration of software. It is necessary to place greater focus on integrating embedded software into a fully functional system.

Using abstraction and virtualization. Generating software from higher-level models can improve maintainability and decrease programming errors, while also improving development speed.

Resolving legacy. So-called legacy software and systems still constitute most of the software running in the world today. New software developed with novel paradigms and new tools will not run in isolation, but rather have to increasingly be used in ecosystems of connected hardware and software, including legacy systems. There are two main areas for innovation here. First, we need to develop efficient ways of improving interoperability between new and old. Second, we must innovate how to (incrementally) migrate, rejuvenate, redevelop and redeploy legacy software, both in isolation and as part of a larger system.

Lifecycle management of software. Industrial systems have a long lifetime, often up to 30 years. If software is not effectively maintained, the software becomes overly complex until it is no longer sustainable. Practical challenges that require significant research in software sustainability include: (i) organizations losing control over software; (ii) difficulty in coping with modern software's continuous and unpredictable changes; and (iii) dependency of software sustainability on factors that are not purely technical.

Digital platforms. It is fair to assume that most future software applications will be developed to function as a part of a certain platform, and not as standalone components. In some domains, this idea has been a reality for a decade (e.g., in the AUTomotive Open System ARchitecture (AUTOSAR) partnership, which was formed in 2003). However, guaranteeing quality properties of the software system (e.g. in safety and security) is a challenging task, and one that only becomes more complex as the size of software applications grows. - Europe is facing a great challenge in the lack of platforms that are able to adopt applications developed by individual providers into an ecosystem. Or, to put it the other way around, there are several (too many) platforms emerging, both from industrial software vendors or departments and from series of big public funded projects.

3.4 Quality, Reliability, Safety, Cybersecurity and Trust

Increasingly, industrial technologies are being regarded as critical applications by law, meaning that **extensive validation, verification, testing and licensing**

procedures must be in place. Security must also be embedded in all engineering tools, which strongly reminds that safety is not achieved by testing alone, but should be **built in** or integrated into every lifecycle stage. Security and cybersecurity are the other side of the coin in the distributed, remote or networked applications that contemporary communication technologies effectively employ. **Lacking useful security could easily be a showstopper**.

Since safety or security is difficult to achieve and prove, industries prefer to talk about **trust** and how they expect (and assume) safety and security will be in place for their business partners. In short, there must be no nasty surprises between trusted partners in terms of security issues.

As regards **privacy**, there is much idealistic urging by researchers, software enthusiasts, etc, for open data and open software. However, certain data must be kept private by law. In addition, critical applications have be sealed and protected once they have been finalized, otherwise their safety, security, functionalities, etc., cannot be guaranteed. Most industrial applications also involve a great deal of engineering effort and creativity, are very extensive and constitute the core asset of companies that must be protected. Competitive business situations could therefore result in a cautious attitude towards open data and software. Nonetheless, industries sometimes do not entirely know what data, etc., it is beneficial to keep private and what should be open. In the era of AI, it may be a challenge to know in advance what could be discovered, for example, in the vast amount of factory or machine data. It is better be safe than sorry! Open interfaces, standards, etc., are good examples and forms of practical openness.

The realm of critical applications has been expanding over the years. Application sectors that have been considered critical include: nuclear industry, certain chemical and petrochemical industries, medical and food, health-care instruments, moving machines, transportation, aviation, and space. The list has been expanding over the year and is expected to grow, since more and more applications will be dealt with law and authorities. At the same time, quality in general has become a necessity in practice and, therefore, techniques and procedures developed for critical applications become relevant wider.

Risk, dependability, and reliability were originally developed for mechanical or electrical system and the core concepts were gradually adapted to software which proved to be most challenging. Software engineers were not accustomed to regard their modules, etc., as critical that had to be ascertained and proven, too. The expanding use of software, however, did not give a relief, and nobody can claim today that the issue is totally solved. Good progress has been made, though.

In the past, critical software should always be small and simple so that the required reliability could be achieved and proven. Also, the hardware underneath should be ascertained. Critical systems must be kept isolated from less critical parts. When entering the extensive, complex, highly communicative systems, with criticality characteristics, the following challenges and opportunities are foreseen:

- Dependable connected software architectures
    - Software reliability in the face of infrastructure instability.
    - Dependable edge and cloud computing, including dependable and reliable AI/ML methods and algorithms.
    - Dependable communication methods, protocols and infrastructure.
    - Formal verification of protocols and mechanisms, including those using AI/ML.
    - Monitoring, detection and mitigation of security issues on communication protocols.
    - Quantum key distribution ("quantum cryptography").
    - Increasing software quality by AI-assisted development and testing methods.
    - Infrastructure resilience and adaptability to new threats.
    - Secure and reliable over the air (OTA) updates.
    - Using AI for autonomy, network behavior and self-adaptivity.
    - Dependable integration platforms.
    - Dependable cooperation of system-of-system (SoS).

- Dependable softwarisation and virtualization technologies
  Changing or updating software by retaining existing hardware is quite common in industrial domains. However, keeping existing reliable software and changing the underlying hardware is difficult. By decoupling software functionalities from the underlying hardware, softwarisation and virtualization are two disruptive paradigms that can bring enormous flexibility.

- Combined SW/HW test strategies

- Trustworthiness
    - Ensuring cyber-security of systems, including AI.
    - Defining different methods and techniques of trust for a system, and proving compliance to a security standard via certification schemes.

ISBN: 13 978-952-5183-59-7

- Defining a method for multiple standards via the composition of certified parts.
- Enabling developers to have a flexible means to demonstrate security capabilities.
- Developing technologies, methods and techniques to ensure cyber-security at all levels.
- Definition and future consolidation of a framework providing guidelines, good practices and standards oriented to trust.

- Safety and resilience of (autonomous AI) systems in dynamic environments
- Modular certification of trustable systems and liability
- Dynamic adaptation and configuration, self-repair capabilities, (decentralised instrumentation and control for) resilience of complex and heterogeneous systems
- Safety aspects related to the human/system interaction

In Finland, the Finnish Automation Society published the book *Automaation tietoturva* (Cybersecurity in Automation), in 2005. **The second and a very much updated version** has been edited during 2020-21, and the volume will be published at this *Automaatiopäivät24* event. We very much recommend the audience to acquire the book and learn.

3.5 Digital platforms

As noticed earlier, most **future software applications will be developed as a part of a certain digital platform**. One cannot do without them. A digital platform is an operating system like complex software entity by which application pieces of software, or modules, or objects can communicate, work together, etc. A platform offers a means for lower-level data and control signalling, tools and means for compound application composition and configuration, managing cybersecurity, i.e., all extra software engineering one would need to build extensive applications like factory or enterprise systems. Some platforms are intended to end-to-end engineering and business processes, whereas, some others are dedicated to modelling and simulation, maintenance management, etc.

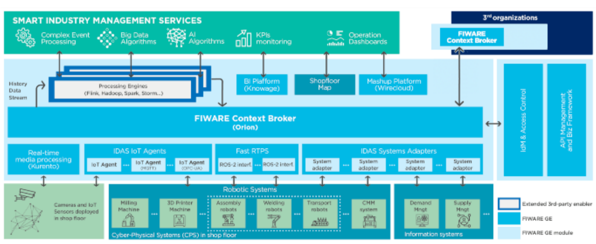


Fig. 15. FIWARE digital platform for manufacturing.

Many software houses build their offering on top of a growing digital platform. Platforms are also sold as a backbones for customer's expanding digital applications' portfolio. The advent of Industrie4.0 gave a model for most of the digital platforms for manufacturing or production, and throughout the 201X'ies, they all started to converge towards Industry4.0 definitions (platforms are big softwares so changes do not happen immediately).

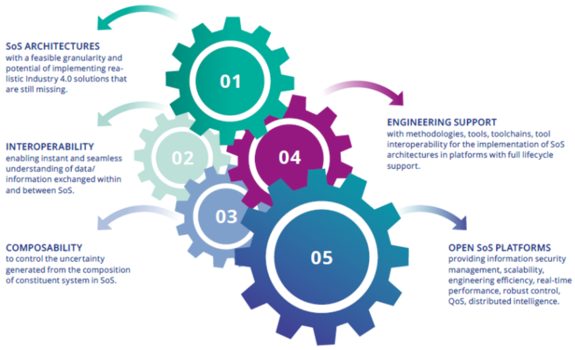


Fig. 16. Evolution of digital frameworks in EU funded projects.

There has been a number of EU funded big projects, or series of projects, where several digital manufacturing platforms have been evolving. Project acronyms such as SOFIA, ME3GAS, IoE, ARROWHEAD, PRODUCTIVE4.0, AFarCloud, and FIWARE are famous, they have spent 100M€+ EU funds each, and reached on the order of 1000 European companies, universities, etc. The major European software and engineering companies are developing their own platforms (Siemens, ABB, Schneider – as examples of the large). US, Japan, China, Korea have their own significant initiatives.

Perhaps five years ago, European Commission declared its Digital Single Market vision [8], meaning every company or stakeholder, big and small, should have a viable opportunity and means to connect to any manufacturing or industrial application or project, as part of the growing European ecosystem. The digital platform landscape is still very fragmented, with open and closed, vertical and horizontal platforms in different developments stages and for various applications. There is a strong need for interoperability / standardisation and orchestration / federation of platforms. The trend towards agile, composable, plug and play platforms and more decentralised, dynamic platforms supporting AI at the edge, needs to be encouraged and supported. With particular emphasis to be given to practical accommodation of SME uptake and deployment.

Like standardization (also an important part of platforms), **there are definitely too many platform**

developments ongoing. Perhaps there are too many egos who want to differentiate. Automation has also been a sector where so-to-say **semi big companies dominate**, i.e., big in a big country or big in a subsector – but not superiorly big globally or even in Europe, at least in contrast with certain global IT sector players (Microsoft, Google, Amazon, etc.). Industry applications are also extensive and complex, differ from sector to sector, and therefore one size does not fit all. Even the EU projects seem to concentrate in a certain way. Some are Mediterranean bound, some are Northern bound, or there are just clusters of old acquaintances among partners. The projects are also up limited of partners.

Existing gaps can still be found in the following topics:

- Moving the focus to industrial and engineering applications. It is important to win the global platform game in various application sectors (which are strong today), and to effectively develop high-level outperforming applications and systems for actual industrial and business requirements.
- Preparing for the coming 5G era in communications technology, especially its manufacturing and engineering dimension.
- Long-range communication technologies optimized for machine-to-machine (M2M) communication and the large numbers of devices – low bit rates are key elements in smart farming, for instance.
- Solving the IoT cybersecurity and safety problems, attestation and security-by-design. Only safe, secure and trusted platforms will survive in industry.
- Next-generation IoT devices with higher levels of integration, low power consumption, more embedded functionalities (including AI capabilities) and at a lower cost.
- Interoperability-by-design at component, semantic and application levels.
- IoT configuration and orchestration management allowing for (semi)autonomous deployment and operation of large numbers of devices.
- Decision support for AI, modelling and analytics, in the cloud but also in edge/fog settings.

3.6 Edge and cloud computing, 5G

The computing platforms have been under significant transition during recent years, and this will probably break the prevailing factory information and automation system architectures we have had since 1980'ies, cf. above. Thanks to the intensive autonomous driving R&D of car industries, these schemes are also penetrating to machine industries. It is now envisioned that already near future (piloted today) system architectures consists of three layers of computing devices, see Fig. 17. 1) **Embedded computing** reside very close or attached to the machinery or process. They can be seen as a legacy of the near past IoT devices connecting directly to hardware, process material, etc., performing some embedded intelligence and communicating to and from the here called Near Computing nodes (computers), either wirelessly or wired. These **Near Computing** devices (2) are often called **edge computers**, routers, or local servers. Near computer nodes are powerful computers themselves and communicate both to the embedded computing cards and to the **cloud** (3) via **internet**. Obviously, there will always be needs for many kinds of more centralized computing, where the so-called cloud computing has been a kind of mainstream, at least on the R&D frontiers. But many kinds of central servers, etc., fit easily into the concept. – The term "edge" got its name doe to the fact that it refers to something at the edge of the internet or cloud. Some people propose to place a layer of special computers to coordinate, etc., the edges. They are called "fog".

A special boost in the picture comes from the **5G** technology. There, the 5G communication technology takes care of the edge-to-cloud and edge-to-edge communication which will be faster than anything before. The 5G base stations are no longer mere antenna poles but equipped with very powerful computing nodes, and the densely installed 5G base stations will take the role of edge computers.

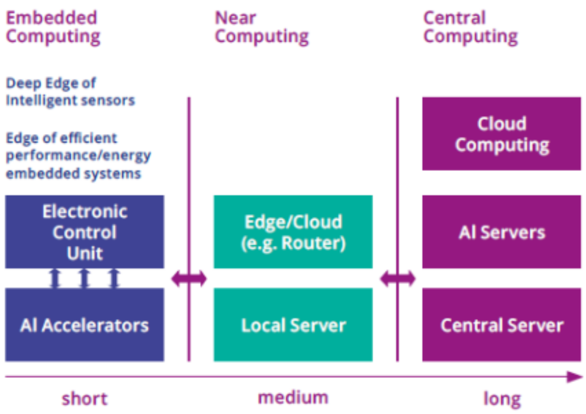


Fig. 17. Emerging computing architecture of embedded computing, near computing and central computing. [6]

Such new computing architectures are now being piloted in various applications, including autonomous driving or bringing advisory or pleasure features to vehicles, smart buildings, pilot factories, etc. Advanced condition monitoring and management, telepresence, and other remote service business applications are surely finding these technologies interesting.

Although touted as a revolutionary technology, this concept seems to appeal to industry. Looking from distance, it actually does not look very different!

Automation systems have always had local and short cycle-time computations, line and factory level computations, and then central or remote computations. No need to change the overall mindsets. The computing and communicating capacity seem almost unlimited and appealing to all latest enabling computing and communication intensive technologies discussed in this paper.

Responsible and experienced industrial engineers do have concerns and questions here:

- Critical automation applications have been implemented on dedicated PLC and DCS hardware for decades, with dedicated design methods, tools, and languages. A lot of engineering experience is embedded in these. Do they easily port to the new paradigm? What needs to be ported? What comes extra or instead? Are the new devices and softwares reliable enough? Can I keep my assets of previous designs?
- Is wireless communication safe and secure? Or if and as we want to keep the old and proven solutions, can we operate stand-alone if the 5G connection is missing?
- Industrial systems have anyhow long lifetimes, often 30 years, and their automation is typically renewed every 10-15 years.

Despite the above cynics, it is better to stay open-minded and look forward to the opportunities of this partly new paradigm.

3.7 Responsive and smart production

The ability of quick reaction to changes, efficient work allocation, predictable quality, fair circulation of tasks, and the ability to tailor tasks for individuals are important targets for a responsive and smart production.

Agile factories and networks are required to deliver products with a high degree of personalization (high-mix low-volume, lot size one, highly customized manufacturing), Therefore, to make a manufacturing plant more constructible, adaptable, and dependable, improved, scalable approach to adaptivity and resilience is needed. This approach must be general, cohesive and consistent across all interconnected lifecycles of the system components. Many European initiatives and reports cover this topic:

1. EFFRA calls for “Agile value networks design, manufacture and deliver innovative products with a high degree of personalisation” [9]
2. SMART-EUREKA mentions “Intelligent and adaptive manufacturing systems” as a main research and innovation domain. [10]
3. ManuFuture: “Factories will adapt and become resilient to foreseen and unforeseen changes in the market and in technology.” [11]
4. World Manufacturing report 2018 indicates that rapidly responsive manufacturing is a disruptive trend, to “…react quickly to and take advantage of changes in market conditions, customer preferences, manufacturing conditions and social demands.” [12]

Attention for further development is still needed as follows:

1. **Robust optimal production, scalable first-time-right production.** Manufacturing should become more robust, even when facing disturbances. This will require advances in, eg., self-healing and redundant automation systems, first-time-right, zero-defect manufacturing, and predictive maintenance.
2. **Mass customization, personalized manufacturing. Customer-driven manufacturing. Mastering complexity of products, processes and systems.** Progress towards lot-size-one manufacturing and personalized product design will continue.
3. **Resilient and adaptive production, including supply chains. Shortening supply chains. Modular factories.** Resilience is a critical property for a system that can absorb change and adjust its functional organization and performance, so that it can maintain the operations necessary to accomplish its mission in varying conditions.
4. **Cognitive production.** Cognitive production looks to deploy both natural and artificial cognition to enable new analyses and learning that can enable responsive and sustainable adaptable production.
5. **Manufacturing as a service.**
6. **Standardization.** One cannot easily beat the legacies of many standards and their installed base. Instead, focus on bridging systems of diverse standards. Develop semantic technologies as a vehicle to master diverse and multitude of standards. Develop middleware or translator software, or platforms that enable effective connectivity, on high application level. Develop respective digital test and development environments. Develop licencing. Ensure wide acceptance and support of software vendors, engineering offices, end-users.

3.8 Sustainable production

Nearly 200 countries have committed to the Paris Agreement on climate change to limit global warming to below 2ºC [13]. Rapid transformation of all sectors is required. Many European countries have set even more ambitious targets. Digitalization could have a great bearing in reducing environmental impact.

Proceedings - Seminaarijulkaisut
ISBN: 13 978-952-5183-59-7

The vision of SPIRE (Sustainable Process Industry through Resource and Energy Efficiency, [7]) decomposes the above high-level goals into more concrete actions, as follows:

1. Use energy and resources more efficiently within the existing installed base. Reduce or prevent waste.
2. Re-use waste streams and energy within and between different sectors, including recovery, recycling and re-use of post-consumer waste.
3. Replace current feedstock by integrating novel and renewable feedstock (such as bio-based) to reduce fossil, feedstock and mineral raw material dependency while reducing the CO2 footprint of processes or increasing the efficiency of primary feedstock. Replace current inefficient processes for more energy- and resource-efficient processes when sustainability analysis confirms the benefits.
4. Reinvent materials and products to achieve a significantly increased impact on resource and energy efficiency over the value chain.

The ManuFUTURE 2030 Vision [11] puts the intended combination of visions together as depicted below.

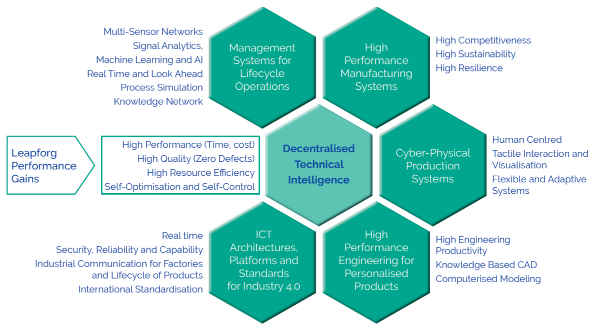


Fig. 18. The visionary manufacturing system for adding value over the life cycle with decentralised technical intelligence (from [11])

Directions for future research and development are:

1. **Life-cycle assessment.** Life-cycle assessment (LCA) is a prerequisite for holistic environmental evaluation, and is a simple but systematic method. However, it requires extensive experience and comprehensive models and data. In practice, mixed combinations often need to be employed – for example, missing measurements must be compensated for by models or standard data. LCA software must also be better integrated into other automation systems.
2. **Monitoring flows of energy, materials, and waste.** It is already commonplace in many sectors (e.g. food, medicine) that material and energy streams need to be completely **traced** back to their starting point. Flows need to be monitored. Sustainable manufacturing needs comprehensive environmental and other **measurements** that may not at all be in place when a particular manufacturing or production was initiated. On the other hand, this is a very typical application of many kinds of IoT sensor and system that can be informed by a careful LCA. - **Discharges or losses mostly happen when production does not occur as planned**, as is optimal, due to mistakes, bad condition of machinery, unskilled operation, and so on. Human factor causes most of the variation in running continuous processes. There should be focus on how to AI assistant or AI optimizer could be used to help operators by giving advices and preventing not optimal changes.
3. **Human-centered manufacturing.** A higher level of formal training may be required for workers in production and maintenance. Greater specialization is constantly introducing product, process or company-specific extra training.
4. **Green Deal.** Policy initiatives aimed at putting Europe on track to reach net-zero global warming emissions by 2050 are put forward in the European Commission's European Green Deal [14]. The Commission plans to review every EU law and regulation in order to align them with the new climate goals.

3.9 Autonomous systems, robotics

The most famous fully autonomy target today is **autonomous driving**. All major car manufacturers have extensive projects. Progress is defined to advance by levels of autonomy, that is we are going to have 1-5 levels of autonomous vehicles, where level 5 is the 100% autonomy, which seems to be years ahead. But the new car models do have more and more driver assistant features that are nice, too.

The autonomy target has also spread to working machines. We already have autonomous mines and harbors, i.e., where the work site is fairly structured. Whereas, autonomous forest machines may remain years ahead.

There is a long-lasting trend **towards higher degrees of automation** level in factories. Actually, hydro power plants have been in practice fully automatic for decades. There are also pilot factories or concepts of fully autonomy. A more common trend is to have fully autonomous or automatic unit processes, production lines, etc. Again, car manufacturers present eagerly fully robotized production cells or lines. Yet, various degree of **semi-autonomous machines and plants** should be regarded as a realistic vision. Actually, recent studies emphasize a more effective **human-machine or human-robot collaboration**.

Machines are more precise and efficient than humans. Thus, replacing or aiding work processes susceptible to human errors, quality defects and safety issues will have an impact on quality and redundant waste. In some industries, such as construction, aerospace, automotive, critical infrastructures complex systems, and utilities, quality issues and the prevention of hidden defects in structures and/or any mechanical and electrical components are extremely serious. Therefore, substantial losses in terms of legal aspects, safety, potential stresses under critical situations of vehicles and aircrafts, substandard end products and quality costs in general - are potential.

Main aims and evolution trends of robots and autonomous systems in the digital industry are oriented toward 1) Production efficiency, speed and reduced costs, 2) Higher precision and quality, and 3) Safety in the working conditions.

However, between the two extremes of entirely manual or fully autonomous, there will probably always be a large area of semi-autonomous equipment, units, machines, vehicles, lines, factories and sites that are worth keeping somewhat below 100% autonomous or digitized. The reasons for this include:

- a fully autonomous solution may simply be (technically) next to **impossible** to design, implement and test
- if achievable, it may be too **expensive** to be realized
- a fully autonomous solution may be too **complex, brittle, unstable, unsafe**, etc.
- a less-demanding **semi-automatic** solution may be easier to realize to a fully satisfactory level

When automation and digitalization degrees are gradually, reasonably and professionally increased, often portion by portion, they may bring proportionally significant competitive advantages and savings that strengthen the position of digital industries overall. However, since automation or digitalization degrees remain well below 100%, the negative effects to employment are either negligible or non-existent. On the contrary, an increased market position could increase the need for more people in the respective businesses.

3.10 Industrial service business, lifecycles, remote operations and teleoperation

The volume and value of industrial services are increasing by 5-10% each year. The share of services has exceeded the share of machinery for many machine, system, and service vendors, not just for a final assembly factory but also for companies in supply chains. Companies are willing to take larger shares of their customers' businesses, at first as a spare part supplier, later about remote condition monitoring, and extending to a number of other tasks that were considered in the past as customer core businesses. From customer point of view, such a shift in business models is called outsourcing.

Especially, when businesses have become global, some services are provided locally, at or close customer sites, while other services are provided centrally from the original vendor or companies specialized to services. Similarly, as there may be extensive supply chains behind the vendor companies, the respective services may extend to supply chain companies. The industrial era of the past is becoming a service era, enabled by high-end ECT/KDT technologies. This distributed setting is also conveniently fitting to modern edge-cloud architectures as computing and communicating platform.

The importance of service business in the future is evident as the service business enables a revenue flow also after the traditional product sales and, more importantly, the service business is typically much more profitable than the product sales itself.

The service business markets is becoming more and more challenging, while the high-income countries are focusing on high-skilled pre-production and on life-cycle stages. Fortunately, in the global service business market, Europe can differentiate by using its strengths: highly skilled workforce, deep technology knowledge and proven ICT capabilities, but the success needs new innovations and industry level changes.

Key focus areas for future development include:

- **Collaborative product-service engineering, life-cycle engineering**. Extending R&D to take into account that products and systems will be integrated to the industrial service programme of the company. Possibly obtaining knowledge to give services for other similar products (competitors!) than own installed base.
- **Training and simulation**. Complex products like aircrafts, drones, moving machines and any teleoperated machineries need simulation environment for a proper training of the human driver/operator.
- **Condition monitoring, condition-based maintenance, anomaly detection**. Performance monitoring, prediction, management. The traditional service business sector, still encountering major challenges in practice. An extension to the above, as targets of services are

ISBN: 13 978-952-5183-59-7

expanded to other topics in customer businesses than spare parts or condition monitoring.

- **Remote engineering and operations, telepresence**. Operating or assisting in operations of industrial systems from remote sites.
- **Decision and operations support**. In most cases decision making is not automatic, whereas, is based on remote expert assistance or extensive diagnosing (AI based, etc.), engineering, and knowledge management systems.
- **Local and global services**. Organizing services locally close to customers and centrally at vendors' site.
- **Fleet-management**. Benefitting of the number of sold items. Obtaining knowledge and experience from number of similar components and machines in similar or different conditions.
- **Edge-cloud solutions**. Implementing distributed service applications on effective edge-cloud systems.
- **Full life cycle tutoring**. Monitoring activities, level of stress and performance-oriented behavior during the all product life to anticipate its end of life, to properly handle its wastes and recycling and for a better redesign of next product generations.

4 EU research instruments and initiatives

4.1 About EU projects

From applied reasearch point of view European researc looks like Fig. x. In the middle, we have European Commission (EC). Perhaps the most well-know part of it is European Parlament deciding ultimately on all major things, including the so-called framework programs that define about all matters of EU funded research. Very soon, EC is expected to decide on the framework programme period of 2021-27, to be called "Horizon Europe", or HEU. But EC is much more than the parlament. It is also a huge administration entity of thousands of officials, some (or quite many) of them working on the details of the framework programmes. This information can be found, starting in https://cordis.europa.eu. Such offices correspond very much to ministries or other organizations in ordinary countries, though often bigger.

European research is organizad by projects. What kinds of projects are possible can be read in the so-called Work Programmes that list and describe the project call texts. Project calls also inform about when to apply projects (dead lines), what type and size of projects are wanted, how much money is available, etc. Then consortia of European partners prepare their proposals to be submitted to EC. The competition for accepted proposals is tough. Project duration is typically 3 yr. Usually, project calls are known 2-3 yr in advance. EC funds 30-70% of project expences, depending on company size and type, project risk, etc.

Besides companies, universities and research institutes can be partners. There must be partners from at least three EU countries per consortium. Very typically proposals consist of groups of subprojects or application cases of 2-6 partners, often coming from the same country. In other words, planning for a proposal, from Finnish point of view, means discussing among few partners about developing some technologies and demontrating them in industrial setting, during that project.

Perhaps this all sounds simple. Also, the above mentioned EC portal guides the readers in somewhat naive tone. But since competition is fierce, in practice identifying proper project calls and editing a winning proposal is as serious and professional work as any multimillion business offer. Having experienced partners onboard is essential. There are also consultancy companies that are specialized to proposal writing. Doing project work is very much the same as in national joint projects with public funding involved, even including reporting, reviews, etc. In Finland, VTT is the largest beneficary of EU funding, followed by some bigger universities and some R&D intensive companies.

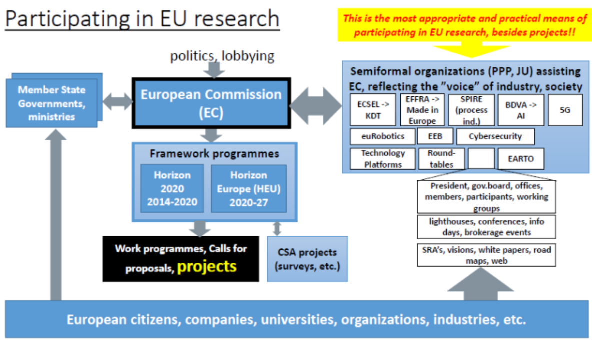


Fig. 19. From needs to running projects in EU.

4.2 Affecting EU research policy

EC or its due offices is in the hart of deciding on work programmes that define the calls and projects. In principle, there is the official channel through MEP's and memberstates' parliaments or governments and administration.

Another channel exists through many organizations of various topics, purpose, background group, etc. For instance, Orgalim (https://orgalim.eu) represents 770.000 technology industry companies. EARTO (European Association of Research & Technology Organizations, https://www.earto.eu) represents VTT and other national research institutes. There are discussion and strategy paper writing forums called

Technology Platforms. They are usually partly EC funded (and partly by membership fees, etc.). In many ways, they gather, edit and communicate the ”voice” of actual industry and research, needs and possibilities of scinece, etc., to the officials of EC. Anybody or any company, etc., can usually participate but expertise and experience of the particular field or technology domain is expected, at least if you want to be heard.

The most important type of such organizations is PPP (public-private-partnership) or JU (joint undertaking). Despite the odd names, they are like European SHOK's we used to have in Finland. They are often formal companies, they have small offices in Brussels, and typically few hundred members (with membership fees) each. They have boards, a president/chairman, governments, small staff, etc., like companies. They organize professional conferences and events, they also have working groups, they write strategies, position papers, or white papers, they organize consortia building fairs prior calls – and they communicate directly to the EC officials about the future project calls. Their opinions, statements, etc., are often clearly visible in the call texts. Unless you are a TOP20 European enterprise, this is the most practical way to affect.

One cannot affect if one does not participate. The membership fees are usually minimal, or you can participate freely in many ways as an attendee. Conferences may have fees for dinners, etc. So, take part in selected activities. Find the best partners in those events. You have to pay, or your company pays, for your travels, work time, etc., its a kind of volunteer work. VTT is there, so are some universities, and some of the most active companies.

Regarding the audience of this presentation, the most relevant PPP's or JU's are:

- EFFRA (European Factories of the Future Research Association, https://www.effra.eu), discrete manufacturing
- SPIRE (Sustainable Process Industry through Perource and Energy Efficiency, https://www.spire2030.eu)
- ECSEL (Electronic Components and Systems for European Leadership, https://www.ecsel.eu)
- ARTEMIS (Industry Association is the association for actors in Embedded & Cyber-Physical Systems, https://artemis-ia.eu), nowadays a section in ECSEL
- BDVA (Big Data Value Association, htpps://www.bdva.eu)
- 5G (The 5G Infrastructure Public Private Partnership (https://5g-ppp.eu)
- euRobotics (https://www.eu-Robotics.net)
- Cybersecurity (European Cyber Security Organisation, https://www.ecs-org.eu/cppp)

EFFRA and SPIRE are clearly application sector oriented PPPs, and there automation or digitalization plays an assisting or enabling role. The other PPPs/JUs are technology focused organizations, but they also try or want to describe how their particular technologies are being used in application sectors. Industry or manufacturing is always the biggest or one of the biggest sectors. The situation is on one hand good, meaning digitalization topics appear a lot on European arenas but, on the other hand, the situation may appear confusing or diverse, and tedious is you want to impact on R&D policies.

The author of this paper has participated actively in ECSEL, EFFRA, SPIRE and BDVA. The contents of this paper is composed and adapted based on the author's contributions, etc., in the SRA's and other core documents. Participating, prior these EU arenas, in several corresponding national, VTT or SHOK programmes gave me a strong background for European circles. A great adventure altogether!

Years 2020/21 have been a transition period between the Horizon 2020 and Horizon Europe programmes, or actually planning and decision making on HEU. There will be some changes in the PPPs/JUs:

- EFFRA will be MadeInEurope
- ECSEL will be KDT (Key Digital Technologies)
- BDVA and euRobotics merge, the new one is AI PPP

These changes are 99% certain. However, as politicians say, nothing is decided until everything is decided. [18]

5 Discussion

5.1 Discussion

Digital automation started with sensors & actuators, PLC and DCS computers, to do continuous or discrete control, to configure alarms, to start machinery, etc. These are still a common practice with factories and machines. As pointed out in Fig. 3 dating back to 2003, or so, we were already visioning and sometimes already piloting many other factory or machine management functions, including machine learning (which became the AI of today), modelling and simulation, remote measurements, and MES level systems. Connectivity was already at that time understood to become a major challenge.

Fieldbuses came, internet and the wireless became means for communication, PC's, PADs, and smart phones emerged. More and more business functions became digitized and modules needed to be more tightly connected. So-called hyper-connectivity is regarded today as one of the source of complexity.

In 2000, modeling and simulation was to a large extent a laboratory excercise strengthened by diverse industrial pilots. The same was true for AI methods, image processing, model-prodictive control, .ect., a lot science behind but few real industrial break throughs. Today and in the foreseen future, these methods and many more, have become an industrial reality. Admit though that a lot of potential is still unused. On one hand companies may still hesitate, but on the other hand, researchers often do not understand to what level and maturity do they have to develop their algorithms to make them industry friendly.

Multi-technology is a must. Some people call these cyberphysical systems, i.e., connecting the digital to the physical. This is easier said than done. Nevertheless, **software is the core**, call it a forth or fifth generation of industrialization, or, a digital transformation. So make yourself a giant in software engineering. Otherwise, you can not implement AI, M&S, digital frameworks, etc., in your businesses.

5.2 Apologies

I have written this piece of text by leaning on several strategy paper, white paper, policy paper, etc., documents during my active years regarding EFFFRA, ECSEL, BDVA and SPIRE organizations. Sometimes, I have been a leading contributor, e.g., in the ECSEL SRA – Digital Industry chapter, or in the EFFRA ConnectedFactories2 proposal text, whereas, otherwise I have been a major or minor contributor for each particular text. I am confessing here now that I have adapted or copied text from those earlier documents more than would have been appropriate for any serious scientific paper. My humble excuse is, however, that I have adapted here mostly my own text from those documents. Since all documents have been teamwork, there are occasions where I for sure know that I borrowed text from European colleagues, and unfortunately in some cases the original contributor has become unclear to me. I do trust my fellow editors from the past for their understanding.

When editing those earlier documents, we always had limitations, e.g., a page limit or maximum number of selected R&D priorities for a strategy, or the target of some particular paper was specific. Here I wanted to make a more complete synergy of matters that I consider relevant to this technology scene. The multitude originates from the fact that industry is big and diverse, application sectors vary, and they always want to benefit of all significant technologies the sole digitalization would offer. Finally, I am a senior citizen now, retired from my long-time employer, VTT. Perhaps I do not expect myself discovering anything new.

References

[1] "Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution". Vdi-nachrichten.com (in German). 1 April 2011.

[2] ÄLY - Älykkäät automaatiojärjestelmät 2001-2004: loppuraportti (Intelligent Automation Systems, final report), edited by Olli Ventä. https://data.nationallibrary.fi/bib/me/W01086261400.

[3] Reference Architectural Model Industrie 4.0 (RAMI 4.0), https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf

[4] What is Industry 5.0?, https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en

[5] Industry 5.0, Publication Office of EU, https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/

[6] ECS-SRIA 2021, ECS Strategic Research and Innovation Agenda, Published on 12 Jan 2021. Downloadable eg. in https://artemis-ia.eu/.

[7] SPIRE P4Planet 2050 Roadmap, Dated: 2 March 2020, www.spire2030.eu.

[8] EU Digital Single Market, https://eufordigital.eu/discover-eu/eu-digital-single-market/

[9] EFFRA Vision for a Manufacturing Partnership in Horizon Europe: 2021-2027. https://www.effra.eu/Sites/Default/Files/190312_effra_roadmapmanufacturingppp_eversion.pdf

[10] Eureka Smart Advanced Manufacturing Technology Roadmap, www.smarteureka.com/mediafiles/2/1517507429-20180201-smart-cluster-technology-roadmap-en.pdf.

[11] MANUFUTURE Vision 2030, www.manufuture.org/wpcontent/uploads/Manufuture-Vision-2030_DIGITAL.pdf.

[12] 2018 World Manufacturing Forum Report: Recommendations for the Future of Manu-facturing. World Manufacturing Forum, 2018, www.worldmanufacturingforum.org/report.

[13] Paris Climate Agreement, https://www.worldwildlife.org/pages/paris-climate-agreement

[14] A European Green Deal, https://ec.europa.eu/info/strategy/priorities-

2019-2024/european-green-deal_en
[15] Information of all European Union research, https://cordis.europa.eu/
[16] Shaping Europe's digital future, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en
[17] Software Engineering Handbook – SWEBOOK, https://www.computer.org/education/bodies-of-knowledge/software-engineering/v3
[18] European Commission, Orientations towards the first Strategic Plan for Horizon Europe, December 2019. https://ec.europa.eu/info/files/orientations-towards-first-strategic-plan-horizon-europe_en
[19] International Council of Systems Engineering, INCOSE, https://www.incose.org/