Acta
Technica
Jaurinensis

Vol. 12, No. 1, pp. 1-12, 2019
DOI: 10.14513/actatechjaur.v12.n1.466
Available online at acta.sze.hu

Comparison of Logistics Performance Measurement Tools

D. Strommer[1,], P. Földesi[1]

**[1]Széchenyi István University, Department of Logistics and Forwarding
Egyetem tér 1., H-9026 Győr, Hungary
e-mail: dianastrommer@gmail.com**

Abstract: In today's word the role of logistics is getting more important in the operation of enterprises. The competition is big, and cost is one of the most important factors. Logistics is a field which can highly support the reduction of the costs. From the other perspective – the customer satisfaction – logistics also has the role of a supportive function. To get the most out of these two big pillars logistics operation needs to be monitored and measured to give room for further improvement. Currently several methods are available for performance measurement. In this article we present comparison of four of the mainly used performance measurement tools.

Keywords: logistics performance, performance measurement, comparison

1. Introduction

This report presents four of the most preferably used performance measurement tools of logistics operations. From the cost perspective logistics is getting more into the focus as this is the area which can be improved and as a result cost reduction can be reached. From the customer's point of view, it is also playing important role in connecting customer and the enterprises and making fruitful cooperation.

The network-oriented development of supply chain led to even bigger complexity. High difficulty is coming from the different goals and perspectives of the different supply chain echelons. The maximalization of personal goals is not working anymore due to the high variety of goods and services which is available for the final customers. Due to the big number of competitors in each level of the chain it is getting complicated to succeed. Reaching customer satisfaction, minimising cost of logistics operation and harmonizing performance of echelons of the chain is crucial.

There are several measurement tools available to evaluate logistics operations. Each of them has different advantages, aims and focus. There are lot of case studies available, so learning the usage of the tools is not very difficult. The bigger problem is the choice of the alternative fits the most to the goal of enterprise, supply chain or network.

The aim of this paper is not only briefly presenting the four performance measurement systems but also making a comparison between them. The goal is not ranking the tools as both means proper support of performance evaluation process but comparing them based on defined features. This comparison should help choosing the proper tool based on the features defined.

2. Performance measurement tools

To evaluate supply chain operation and the member of the chain we need to make measurements. These measures should not be simple as they are evaluating a complex system, but they are highly needed to support strategical decision-making process. There are several tools, methods to use and it is hard to point out which is the best. The performance measurement has wide scale. We can have metrics from a single measurement (such as total cost of the full operation) to complex system which takes into consideration several viewpoints. In today's world the complex measurement is preferred as it is collecting several indicators in a group and tries to evaluate them together. Using a single indicator can easily mislead the evaluation mainly if the targeted area is wider than an enterprise [1].

In the following chapter I will introduce four of the most favourably used performance measurement tools in logistics. All of them is complex tool with several viewpoint which evaluates the operation as a complex process.

2.1. Balanced Scorecard (BSC)

The Balanced Scorecard is a frequently used measurement system which is not limited for measuring financial results. The idea was developed by Robert S. Kaplan and David P. Norton in 1992. The tool is mainly supporting the work if the strategical goal is clear and the number of metrics we would like to measure is limited. For structuring the result Balanced Scorecard use four perspectives [7, 8]:

1. Financial perspective
2. Customer perspective
3. Internal business perspective
4. Innovation and growth perspective.

This model can measure operative processes, but the aim is supporting strategy and long-term changes. The main point in Balanced Scorecard is staying flexible, adaptable with keeping the ability to handle the complexity of measures. The main groups of indicators stated in the model are important in all enterprises and the breakdown of the measurement enables the customization of the system even on enterprise level. Besides the company can also decide regarding the weight of the perspectives based on unique preferences. With regards to the perspectives it is also important to find a way to connect them and point out the parts where they can influence each other [3, 7].

Using this method, the two main pillars which was mentioned already can go hand in hand. The cost pressure can be reflected in the financial perspective, but also the customer satisfaction can be taken into consideration. In the model what is more than the already mentioned points are that we can also connect these disciplines to internal business processes and we can see how the development of the internal processes can help with the other factors. What is more, innovation and growth/adaptability can be also integrated into the complex evaluation of the operation, strategy. Basically, the model tries to answer the question ‘How does the company succeed?’ with the non-financial indicators to support the financial goals and targets [9].

We can differentiate between the perspectives based on the observed period also. Financial perspective shows the financial results of the company, the decisions are made, and the changes occurs due to them. In contrast the other segments mainly focus on the future. What are the areas which can impact future result? How can their improvement help from financial point of view? The limited number of measures also concentrate the focus of decision makers on the focus areas and information not need to be selected, searched out from huge datasets [9].

Having not only financial perspective is a big advantage of the model, but we cannot eliminate the importance of the financial results. In the end that numbers are still the easiest to compare and show the most objectively e.g. the cost of the spare capacity won by the changes [6].

2.2. Performance Pyramid

This model has been developed by Lynch and Cross. The model is fitting in in structure of the company hierarchy. The top of the pyramid is matching with the company vision. These goals are supported by the market and financial related indicators from the tactical planning level. In the same level of hierarchy but more into details we can find the elements which have effect on the above-mentioned results such as productivity and customer satisfaction. On the operative level we find the elements that can be influenced the most. These are the followings: quality, waste, delivery and cycle time. The below figure shows how the Performance Pyramid is structured [10].

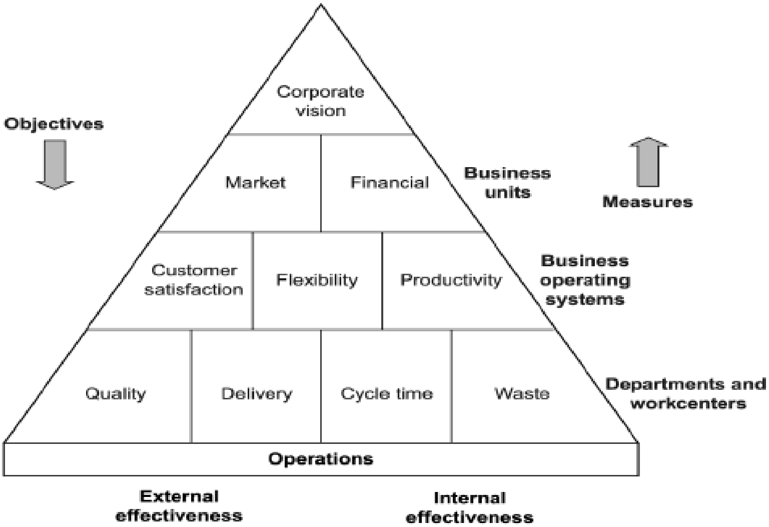


Figure 1.: Performance Pyramid [9]

The core point of the Performance Pyramid is, that instead of top down thinking bottom up method is used. The final goal is placed in the top of the pyramid: corporate vision defined by the company. That makes the basis of lower levels, where the sub-goals are defined based on this. Finally, the steps are initiated in the day-to-day operation. Even though goals are defined strategical level the model is still working in a bottom up way and because of this the goals are defined in all levels. Finally, operative changes turn into strategical goals. This model supports one specific process or problem's evaluation and generation of action plan for it, for continuous improvement it is less helpful. It can be mainly used with traditional, hierarchically shaped companies [9].

2.3. Tableau de Bord

This model is a French performance metrics system also called as French Balanced Scorecard. It used as best practice in several French company's operation. The method is concentrating mainly the control of operative processes. The aim is providing overview and control of the company focusing on the future. The goal is not finding deviations for changes but also direct repair of them [4].

The definition says that Tableau de Bord is a "tool for the top management of the firm, allowing it a global and quick view of its operations and of the state of its environment" [11]. Initially the tool was used for reporting to initiate conversations on fixed items and hierarchical split of tasks. By the time to adapt to changing requirement the model turned to a performance analyser tool. It has three main elements: objectives, action variables and actions. To reach given objectives (goals) the needed elements should be defined and that is covered by the action variable, which is key factor in reaching the given objective. At the end of the day action plan is generated by the action variables. In this process action variable is key, it need to be controllable and if action variable occurs the probability of the objective following should be high. The selection between the objectives what we want to influence should be based on Pareto's rule, so the most significant improvement area should be selected. In this case there are several objectives (areas to improve) continuously know and always the most relevant is checked.

Tableau de Bord can be used in the different hierarchical levels in a bit different format adapting to the special needs. Traditionally objectives are formed in top level and the responsibility is delegated down to lower levels as more detailed knowledge is needed for the next steps. In the end objectives and action variables are connected. For the action plans defined in the process at least one performance indicator need to be defined. Due to the way responsibility is assigned to the right level cross functional coherence is also helping to reach the goals. As I already mentioned BSC and Tableau de Bord is willingly compared to each other because of the strategic approach broken down to action points. Although they differ for example on the concept and structure [2].

2.4. Supply Chain Operations Reference (SCOR)

The model follows process approach aiming to reach effective operation in the supply chain. All activities are checked connected to the material flow and focus on operational efficiency. The model is clearly not following a reporting or analysing approach. It is based on the following steps: plan, source, make, deliver. Plan means the analysis of market information and trends, source stands for the procurement system, make covers the manufacturing, deliver is the process how finished goods reach the customer. It can be completed with and additional element: return, which cover the process of returning goods if needed. It is going beyond a measurement function and aims also to evaluate the issues defined. SCOR model is considering the following attributes: Delivery reliability, flexibility and responsiveness, cost and assets. The focus is on logistical flow and the echelons participating on it. Finance can be part of the measurement, but it will never be in the in the focus area.

The below figure shows that connected SCOR evaluations can give feedback on the operation of the chain. The model can be used in each element but for real improvement the interacting echelons cannot be handled separately [9, 12].

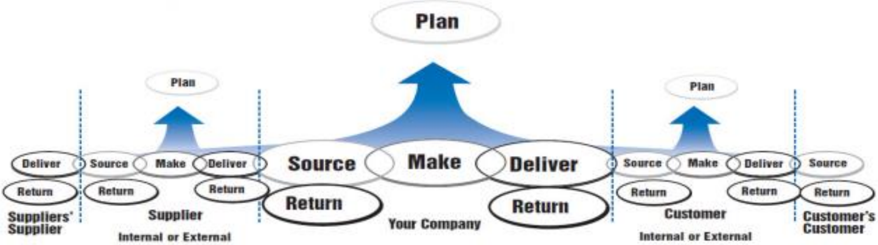


Figure 2.: SCOR model [5]

The model has four levels regarding the implementation. The first level defines the main supply chain processes, support the SCM objectives. The second level goes more into details, explain main process categories. The third level of breakdown consist further benchmarks, information, explanation on processes and capabilities. Finally, the fourth level stands for the implementation [5].

3. Comparison

In this report I summarised the basic information about four performance measurement system. Each of them is widely used in the industry. All of them has advantages and disadvantages. In the following chapter I compare the methods presented before based on the following features:

- **key indicators**. Are the key indicators defined in the given metrics?
- **perspectives**: Are the model defines different view-points, perspectives or categories of measures?
- **non-financial**: Are the measured parts beyond financial attributes?
- **strategy**: what strategy is the model following? bottom up or top down?
- **network**: is the model designed to measure a company? or capable of measuring a network or chain?
- **operations**: is the analytics broken down to operational level?
- **orientation**: customer or company point of view is followed? process or result oriented measures are used?

- **management tool**: is the measurement designed partly or fully to support strategical decision-making process?
- **reporting**: is the aim of the model fully or partially report about the status of the performance?

The comparison is made through different angles. Each of them is important because they are highly differentiated between the companies who use it. The below table shows how the introduced methods can be evaluated based on them.

Table 1. Comparison of performance measurements

Features	***BSC***	***Performance Pyramid***	***Tableau de Bord***	***SCOR***
key indicators	yes	no	yes	yes
perspectives	yes	no	no	yes
non-financial	yes	yes	yes	yes
strategy	top down	bottom up	top down	top down
network	yes	no	yes	yes
operations	yes	yes	yes	yes
orientation	both	process	both	process
management tool	yes	no	yes	no
reporting	yes	no	yes	no

3.1. Methods based on the defined features

Balanced Scorecard is a management tool set up for support strategical goals, decision making of the company. The frame of the model is based on four perspectives which define the structure and ensure that non-financial measures can also take part in the examination, group all the relevant indicators. The action points, next steps are broken down to operative level, but the decision is coming from higher hierarchical point of view. The BSC can be used for a single echelon of the supply chain or for cooperation or network of companies, it depends on the finally defined indicators under the different perspectives. The view-point is also quite flexible, defined by the indicators. It can be both process and result oriented, it is also mainly depending on how the weights of the perspectives set and what kind of indicators are defined during the measurement.

Performance Pyramid is providing company specific support, not useable for networks. The focus is on the development of the operational processes or elimination of a problem with a bottom up strategy. Goals are not predefined, they are set based on the operative level. Instruction, frame or set of indicators are not given. The model is more a process improvement method than a set of metrics. It is based on hierarchy levels. The method is beyond financial measures. The method is not meant to be a strategical decision-making tool but an operational improvement system. Objectives are existing from the top down communication, but the focus is on the bottom up development.

Tableau de Bord as it was mentioned is very similar to the Balanced Scorecard from several extent. It is also a tool supporting strategical decision making and in the same time detailed system with elements broken down to operational level. As it is also defining objectives based on the strategy the strategical flow here also works in the top down way. The objectives mentioned are the key indicators which are chosen to be improved based on the impact for the future, but they are not grouped into categories or put in any frame. Objectives can be financial but do not have to be so this model is also beyond purely financial measures. The method can be easily extended to chain or network of echelons, or it can be company specific. Depending on the scope of the model the action variables and action plans will be different. Regarding orientation the number of possibilities is not limited also. It can be process or problem orientated depending on the content of the objectives and action variables.

The SCOR model is also supporting the strategical decision making but the main goal is not that but gaining advantages of operational changes in the material flow. It is process orientated, so the measures are set to evaluate and improve the processes in place. The direction of indicators is coming in a top down stream of communication. The usage is optimal more for networks than for echelons of the supply chain. The goals are set based on the company vision but the breakdown for tactical and operative level is also part of the methodology. The focus of the measurement is process improvement from different view-points, so it is also beyond the financial indicators. The target of this model is resulting better efficiency in the material flow. It is more of a non-financial approach. The indicators can be grouped in the model-based levels (following the hierarchy) or based on the flow of material in the supply chain (plan, source, make, delivery).

3.2. Differences based on the defined features

One of the main requirement towards the performance measurement tools is being able to handle complexity and support efficiency. This is true for all the examined methods and this is one of the elements what makes them widely used. As among the requirements stated by the users there are no two equals also among the methods there will be no uniform content that makes them adaptable for various scenarios.

Another basic requirement is going beyond the financial metrics and gain better efficiency. This is also part of all the checked methods. They can handle well non-financial metrics. That helps the growth of efficiency and effectiveness of operations.

Among the four tools Performance Pyramid differs the most. All other examined tools are based on key indicators, shaped more in a top down way of thinking. Performance Pyramid aims to evaluate a process or a company it is not willing to handle any wider range.

The SCOR model also a bit differs from all others as it is aiming to execute changes not on one echelon but in a chain or network of companies. Most of the other models are also capable of this but the main structure is not originally designed for this.

Regarding key indicators, measurements in three of the models we can find instructions. In two of these cases the indicators are integrated in a structured way. This structure can help during the understanding or analysis of results or even during shaping other indicators to measure. It can be also used during the decision-making process or reporting of the results. It is also visible that this frame is missing in Performance Pyramid, which is the only tool among the four which is the less compatible with complexity. We can state that if not only processes but also relations and cooperative actions need to be examined than structure of measures can be supportive.

As the initiatives are coming mainly connected to the company's strategy it is not a big surprise that most of the models are working in a top down way. They are fed buy the vision of the company. If we only take this into consideration it may seems that the measures are only for supporting presentation of strategical goals. But the measures are delegated and broken down to operative level. With this at the end of the day top down visions relate to practical steps.

It is also important from what perspective we would like to see the effectiveness of the operation. If we want to have minimum cost and best usage of capacity the focus is in processes. It can be realised for example through setting low stock level or even avoid safety stock. With this no money would stand in stock but it would hardly satisfy the customers need due to the long lead time. It can be managed the other way around as well. That would mean in the given example: safety is set based on agreement with customer or historical sales information. None of the perspectives is good or bad. Focus is question of decision. In this example if the product is hardly replaceable there is no need for the customer or result oriented thinking. In case of e.g. FMCG products it is necessary to start analysis of processes from the required results point.

As table above clearly show SCOR model and Performance Pyramid is set up for process-oriented metrics. It brings the importance of processes and chain or connection of them. While the other two models can be set up from both perspectives. It mainly depends on the key indicators which are phrased in the model. In case on Balanced Scorecard it also depends on the weight we give to each perspective, so we can differentiate.

Both BSC and Tableau de Bord are management tools. That mean that the result of the analysis is preferably used for the decision-making process later. Due to this the models have supportive functionality. In contrast SCOR is aiming to implement changes right away if there is any possibility for them. This does not mean that BSC and Tableau de Bord is not supporting the implementation but means that they have a functionality of management support.

We cannot say that any of the introduced systems are fully dedicated to reporting purpose. This is only a subsidiary function. Two of the examined cases is not created to support any kind of reporting functionality. As it was already mentioned SCOR is a practice-based tool which is highly focusing on the implementation and practice. Performance Pyramid with the bottom up approach is also not a reporting tool.

3.3. Comparison of Tableau de Bord and Balanced Scorecard

As the similarity between the Balanced Scorecard and Tableau de Bord is quite big it also worth to compare only the two them. Both two tools are a top down strategy based supportive methods which aim to examine the operative processes based on the company vision and strategical goals. The measures are not made on high level but translated to day-to-day tasks and with this the vision is directly connected with employee actions.

We can also state that the bypass of financial dominance in measures is aim of each. Both model offers several key indicators what can be used in the frame of the measurement. But it is also recommended in all cases to keep the focus with weights or prioritized KPIs. This ensure that the user would not lose the real results between the huge number of metrics. The mentioned frame is usefully mainly due to the decision-making support functionality of the models and due to the reporting aims.

Although the similarity between the methods are indisputable we still can see some differences. The biggest gap between the model is in the level of predetermination. Balanced Scorecard has four perspectives which strictly determine how the result will look like and partially also determine what measures can be introduced in the model. From this perspective Tableau de Bord is designed freer, not connected tightly with a predesigned structure. On one hand this makes Balanced Scorecard easier to understand and set up but on the other hand it can be hindering factor from the adaptability point of view.

Other difference can be find if the focus or result of the model is examined. In case of Tableau de Bord the aim is defining an objective and through an action plan reach this goal which drives to development. In contrast Balanced Scorecard tries to make measurable, quantitative performance indicators and the focus is more on the result itself than on the development.

4. Conclusion

As the changes of requirements are more and more pushing towards cost reduction the effectiveness of the companies and logistics operation is remaining in focus. Performance measurement tools are supporting this changes and challenges. During the past years they changed according to the modified market needs. In this report four measurement has been introduced. Each method has its own advantages and disadvantages, each of them was created to ensure the effective operation of the company.

This report's aim to give a starting point of selection between the four examined methods based on the user specific requirements. Table 1. summaries the comparison of the different measures. It is clearly visible that in case of network evaluation we cannot use Performance Pyramid, but it is totally fitting with a hierarchical setup. It is also definite that SCOR model is the most practice oriented and focus most on network-based efficiency. Regarding Balanced Scorecard and Tableau de Bord two very similar tool has been introduced with determined structure and defined key indicators. The setting of order is not targeted in this report. Each tool is appropriate in its own field. One of the core tasks is defining the main features of the model we are searching for, so it can support further goals better.

References

[1] N. Asadi, Performance Indicators in Internal Logistic systems, in 2012 International Conference on Innovation and Information Management (ICIIM 2012) IPCSIT vol. 36 (2012) © (2012) IACSIT Press, Singapore, Singapore, 2012, pp. 48-52.
URL http://www.ipcsit.com/vol36/010-ICIIM2012-M0026.pdf

[2] A. Bourguignon, V. Malleret, H. Nørreklit, The American balanced scorecard versus the French tableau de bord: the ideological dimension, Management Accounting Research 15 (2) (2004) 107–134.
https://doi.org/10.1016/j.mar.2003.12.006

[3] P. Chytas, M. Glykas, G. Valiris, A proactive balanced scorecard, International Journal of Information Management 31 (5) (2011) pp. 460–468. https://doi.org/10.1016/j.ijinfomgt.2010.12.007

[4] J. H. Daum, French Tableau de Bord: Better than the Balanced Scorecard?, Der Controlling Berater 7 (2005) pp. 459-502.

[5] G. E. Delipinar, B. Kocaoglu, Using SCOR model to gain competitive advantage: A Literature review, Procedia – Social and Behavioral Sciences 229 pp. 398-406. https://doi.org/10.1016/j.sbspro.2016.07.150

[6] M. J. Epstein, J. F. Manzoni, The Balanced Scorecard and Tableau de Bord: Global Perspective on Translating Strategy into Action, Management Accounting: Official Magazine of Institute of Management Accountants 79 (2) (1997) pp. 28-36. URL https://flora.insead.edu/fichiersti_wp/inseadwp1997/97-82.pdf

[7] Z. T. Kalender, Ö. Vayvay, The Fifth Pillar of the Balanced Scorecard: Sustainability, Procedia – Social and Behavioral Sciences 235 pp. 76-83. https://doi.org/10.1016/j.sbspro.2016.11.027

[8] R. S. Kaplan, D. P. Norton, The Balanced Scorecard - Measures that Drive Performance, Harvard Business Review January-February (1992) pp. 71-79.

[9] G. P. Kurien, M. N. Qureshi, Study of performance measurement practices in supply chain management, International Journal of Business, Management and Social Sciences 2 (4) (2011) pp. 19-34.

[10] R. L. Lynch, K. F. Cross, Measure Up! Yardsticks for continuous improvement, 2nd Edition, Wiley, Cambridge, 1995.

[11] J.-L. Malo, Les tableaux de bord comme signe d'une gestion et d'une comptabilité, in: Melanges en l'honneur du professeur Claude Pérochon, Foucher, Paris, 1995, pp. 357–376, in French.

[12] J. Thakkar, A. Kanda, S. Deshmukh, Supply chain performance measurement framework for small and medium scale enterprises, Benchmarking: An International Journal 16 (5) (2009) pp. 702-723. https://doi.org/10.1108/14635770910987878

Acta
Technica
Jaurinensis

Vol. 12, No. 1, pp. 13-25, 2019
DOI: 10.14513/actatechjaur.v12.n1.488
Available online at acta.sze.hu

Power Semiconductor Trends in Electric Drive Systems

M. Csizmadia[1], M. Kuczmann[1]

**[1]Széchenyi István University, Automation Department,
Egyetem tér 1., H-9026 Győr, Hungary
e-mail: csizmadia.miklos@sze.hu, kuczmann@sze.hu**

Abstract: Nowadays lots of big brands (like Tesla, Nissan, Audi etc.) deal with electric cars and electric drive systems. The first brand deals with only electric drive systems and everyone know this name. These cars are more environmentally friendly, because those operate only with electric energy (this article does not deal with the source of electric energy). The design of electric drive systems is very difficult and complicated task: the electric, thermal and mechanical parameters are very important during the design process. The task is given: it must be designed to reach the most efficient drive system with a low cost. This article deals with the current semiconductor trends and properties, investigates the current electric car drive systems (semiconductor design perspective) and deals with the future trends.

Keywords: *electric cars, power semiconductors, power electric, electric drive system, wide bandgap semiconductor*

1. Overview

Most of electric cars drive system can be divided into the following parts:

- Battery;
- Control electronics;
- Electrical motor;
- Power electronics.

This article is deal with the power electronics, which include (Fig. 1.):

- The three-phase frequency converter, which includes the semiconductors and heat sink;

- The gate drive systems and protection circuits;
- The current and voltage measuring systems;
- The battery management system (BMS) and the battery charger;
- The printed circuit board (PCB) design.

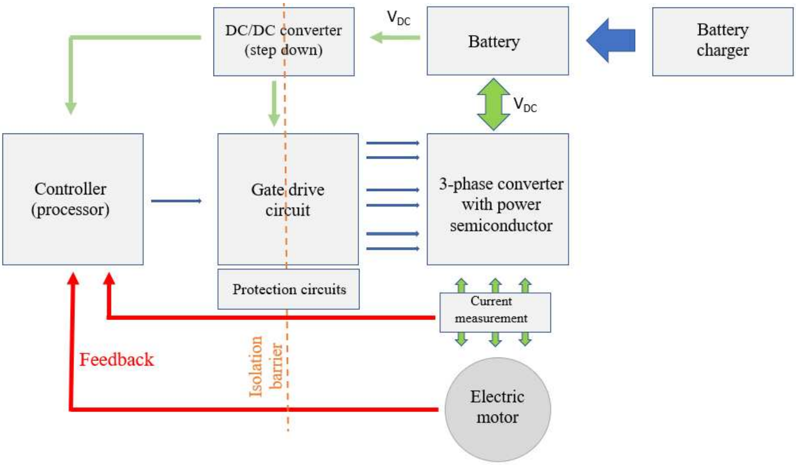


Figure 1. Block diagram of powertrain of electric cars

2. Semiconductors in electric drive system

Semiconductor properties extend to electrical and thermal parameters. The most important electrical parameter is the maximal values, which limit the boundary of operation:

- The breakdown voltage;
- The maximum current, which depends on the frequency and the temperature. In this case it is distinguished normal and surge current (the latest can work in short time);
- Working frequency;
- The switching parameters (capacities and charges);
- and their combined parameters: FBSOA (Forward Biased Safe Operating Area) and RBSOA (Reverse Bias Safe Operating Area).

Nowadays, in the high-power systems are used Insulated-Gate Bipolar Transistors (IGBTs), having high breakdown voltage and high current value, but these parameters are achieved at the cost of switching achievement. A good example is the Tesla Model S: the frequency converter of the car was developed with discrete IGBTs [1].

In the last decade more attention is given to the wide bandgap semiconductors such as the silicon carbide (SiC) and the gallium nitride (GaN). They are called as the “next generation of the semiconductor devices”. These semiconductors are the keys to the high frequency and high efficiency design systems. The high frequency switching includes the lower ripple: However, the lower ripple reduces the EMI disturbances. The high switching frequency means also higher switching losses. Therefore, the choosing of the switching frequency is an important step in the design procedure, for the high system efficiency.

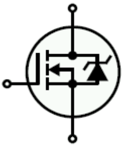
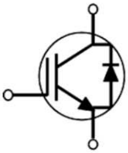
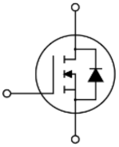
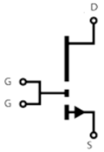
The SiC-based semiconductors have lot of positive advantages: lower power loss, lower on-state resistance, greater bandwidth, which includes the higher switching frequency and the better thermal parameters. Compared to the IGBT the SiC MOSFET in the same switching frequency has lower power loss. First, in the market in Tesla Model 3 SiC-based power electronics are used with integrated gate driver [2] [3].

The common features of GaN-based transistors are the low power loss, the low on-state resistance, the very low input capacitance and the high switching frequency. Also, in case of GaN-based transistors can be found the depletion and the enhancement type. In the practice, the enhancement type is the widespread, because it is off in case of zero gate bias. A new type of GaN based semiconductor has been investigated, called the GaN E-HEMT (GaN Enhancement mode High Electron Mobility Transistor). Basically, it characterizes the same properties, as the SiC-based semiconductors, but it has more positive advantage over the SiC: very high switching frequency (>100MHz), low gate charge and threshold voltage, the structure does not contain "body-diode" (zero reverse recovery loss), very good gate bias properties [3] [4].

The trend shows, that the switching frequency of the system is increasing, therefore the SiC and the GaN based semiconductor will enable innovative improvements in power density, efficiency and costs in the next years.

Table 1. contains the actual (or potentially) applicable semiconductors in electric drive systems nowadays and in the future. The values in the table are based semiconductor manufacturer datasheets. Extremist examples can also be found.

Table 1. Semiconductor properties

Symbol				
	MOSFET	**IGBT**	**SiC-MOSFET**	**GaN E-HEMT**
Breakdown voltage	Medium <1kV	High >1kV	High >1kV	Medium <1kV
Maximum current (pulse)	Medium < 0,5kA	Extreme High >1kA	High < 1kA	Medium < 0,5kA
Switching frequency	Medium ≥ 100kHz	Low <100kHz	Medium ≥ 100kHz	Extreme high > 1Mhz
Switching power	Low <500W	Extreme High >1kW	High >1kW	High > 1kW
Max. operation temperature	High ~150°C	High~150°C	Extreme high* 200°C	High ~150°C
Losses	Middle	Middle	Low	Extreme low
Example	Infineon IPW60R045CPA (600V@230A) 	Infineon FF1200R12IE5P (1200V@ 2400A) 	ROHM BSM600D12P3G001 (1200V@600A) 	GaN Systems GS66516T (650V@144A)

* Depend on the semiconductor casing.

2.1. Casing of power semiconductor

Semiconductors design can be discrete or modular design. In the case of modular design, the semiconductors contain half bridge or three phase bridge. Sometimes the modular case is very useful, because the warming of the semiconductors is the same (if the load is the same and symmetrical). In most cases the modular design contains NTC resistors, which can be able log the thermal behavior. The other advantageous property of the modular design, of the direct liquid cooling (DLC). The high current plugs can be connected with screw and the controller plugs can be connected with soldering. In case of discrete design, the robust casing of semiconductors is recommended (e.g. TO-247): The robust case can be fixed better to the heat sink. Select of casing is depending on the design procedures: For example, in the frequency converter of Tesla Model S discrete IGBT semiconductors have been used, so this solution is not rare in the practice [1].

Another important property is the inductance of casing. At higher switching frequency the parasitic inductance has a bigger impact: the leakage inductance can cause gate ringing in the switching procedures. It can be reduced with several methods, like the distance between the gate driver and semiconductor have to be minimized. If the capacitances and parasitic inductances of the semiconductor devices are small, then the impact of these phenomena is negligible. For example, in case of TO-247 casing this parasitic inductance is about between 5nH and 15nH [6].

In case of the wide bandgap semiconductors the casing must be carefully reconsidered, because such type of semiconductor's working temperature will be higher, than the Si-based one. The requirements of the packaging are the low parasitic inductance, the mechanical robustness and the low losses. Nowadays the surface mount casing in case of high-power semiconductors is not rare. *Table 1.* contains some type of casing.

2.2. Dynamic properties of semiconductor

The dynamic properties of the semiconductors include the switching properties: the capacitances, the charges and the inductances. These properties are defined in the datasheet, like the input, output and transfer capacitance, the gate charge and the parasitic inductance of the semiconductor.

In case of Si-based semiconductors, the input capacitance has very high value, typically a multiple thousand pF. The more capacitance means also more gate charge, which must be moved during the switching procedures. Thus, requires the gate driver more source and sink current, which increase the complexity of the system.

For the good comparability some 650V MOSFET with the same current rating (about 60A, at 25°C) have been investigated. The Si-based MOSFET (STW65N65DM2AG), the SiC-based (SCT3030AL) and the GaN-based (GS66516T) are shown in *Fig. 4.* It has been found that the device input capacitance $C_{iss} = C_{gs} + C_{gd}$ of GaN-based MOSFET is the lowest, the Si-based MOSFET is slightly bigger than SiC. Reverse transfer capacitance $C_{rss} = C_{gd}$ of GaN-HEMT is much smaller than that of SiC devices. Semiconductors output capacitance $C_{oss} = C_{ds}+C_{gd}$ values of the aforementioned devices are similar in case of SiC and GaN. The output capacitance of Si-based is significantly bigger.

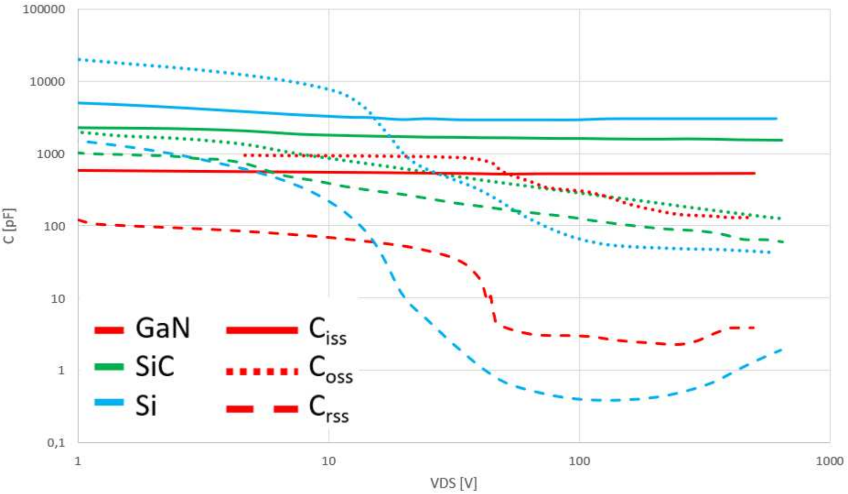


Figure 2. Comparison of Si, SiC and GaN-based semiconductor capacitance

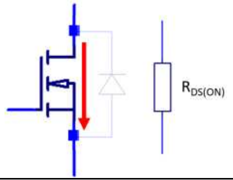
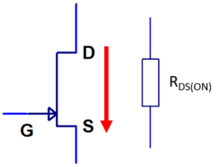
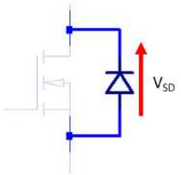
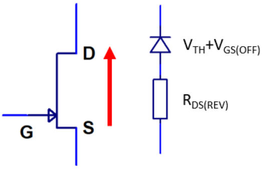
2.3. Body diodes

In electric drive systems the diodes play an important role, because these conduct the current on the load during the semiconductor in off-state, or in synchronous rectifier system. The important properties of these diodes are the forward drop voltage (V_f), the reverse recovery charge (Q_{rr}) and time (t_{rr}). To achieve high switching frequency, low reverse recovery properties are needed. In case of the Si-based semiconductors these parameters are hundreds of ns and µC in magnitude. The same parameters in case of wide bandgap semiconductors are smaller: usually less, then 100ns and the recovery charge is nC in magnitude [7].

In case of GaN E-HEMT the structure does not contain the “classical” body diodes. The properties of the new solution are the same with better switching parameters. Table 2. compares the MOSFET and the GaN E-HEMT semiconductors on and off states, focusing on the equivalent circuit [6].

Not rare the hybrid solution in high power inverters: the Si (body) diode are replaced by SiC Schottky diodes and are used with Si-based IGBTs. The power losses can be reduced significantly.

Table 2. Reverse conduction [6]

Gate state	MOSFET	GaN E-HEMT
ON		
OFF		

2.4. Gate bias voltage

Every semiconductor has a minimum voltage in the control inputs, by which the gate driver have to achieve and have to exceed, for procedures of the safe switch-on. This parameter is the well-known gate threshold voltage. The threshold voltage decreases by increasing temperature therefore the silicon material has lower threshold voltage: the gate threshold voltage has a decreasing tendency in case of new semiconductors. Therefore, the semiconductor gate voltage is a very important parameter. The output parameters of semiconductors, the gate drive losses are depending on this parameter. It is well known, that the gate driver losses are given by [8] [9]:

$$P_{loss\_gate} = Q_g V_{DRV} f_{sw}, \qquad (1)$$

where

Q_g is the total gate charge (or often denoted by Q_{sw});

V_{DRV} is the gate driver voltage;

f_{sw} is the switching frequency.

The gate is fully charged and discharged in each switching period, therefore is very useful, if the value of the gate charge is small. The losses can be decreased, if the gate drive voltage is well-chosen. The basic drive method is the same as that for IGBT's and Si-MOSFET's. The turn on gate voltage is typically between 12V-18V and the turn off voltage is in the most cases negative, approximately -3V to -8V. The maximum value of the gate voltage is typically ±20V. In case of GaN transistors the threshold voltage is lower, therefore the gate drive voltage can be lower in the turn on procedure. The on-off gate voltage hysteresis from -3V to +7V and the maximum ratings are between -20V and +20V. Table 2. summarizes the gate bias level of the mentioned semiconductors [6].

Table 3. Gate bias level of semiconductors

	Si-MOSFET	**IGBT****	**SiC-MOSFET**	**GaN Transistor**
Threshold voltage	1-4V	about 4V	about 3V	1-2V
Maximum rating[*]	±20V	±20V	±25V	-10V/+7V
Typical gate bias value	0V/10-12V	-8V/+15-18V	-4V/+15-20V	0 or -3V/5-6V

* The maximum values are different in case of different manufacturers

** The threshold voltage of high voltage devices is typically larger then for low voltage (e.g. 30V).

3. Semiconductor losses

In the switching procedures the total loss on semiconductor can be divided into two significant parts: the static losses (conducting losses) and the dynamic losses (switching losses). During the switching procedures the gate driver has a power loss, too, which can be estimated by (1).

The static losses depend on just the on-state resistance of the semiconductor and the switching current. In case of MOSFET the on-sate resistance is described by an ohmic resistance. In all other cases it can be described by the linear approximation of the U-I characteristic. The well-known equation in case of MOSFET is a follows [4] [9]:

$$P_{loss\_cond} = R_{DS(on)} I_{D(RMS)}^2. \qquad (2)$$

where

$R_{DS(on)}$ is the semiconductor on-state resistance;

$I_{D(RMS)}$ is the root-mean-square (RMS) current through the MOSFET.

It is very important, that the static losses are independent from the frequency. The on-state resistance of the semiconductors has decreasing tendency, it is shown in *Fig. 3.*

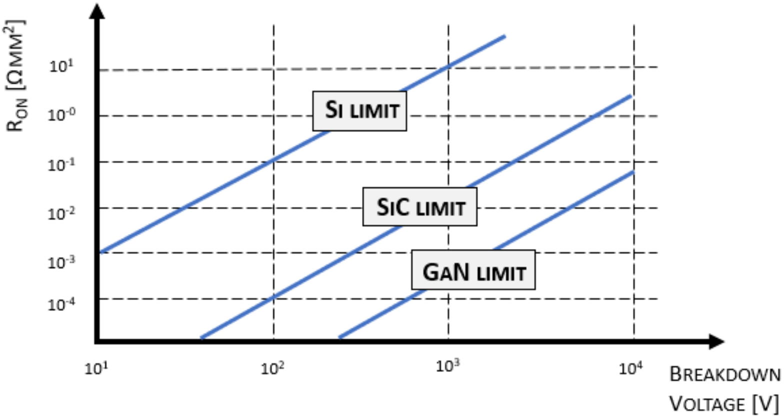


Figure 3. On-state resistance vs. blocking voltage for Si, SiC and GaN [5]

The dynamic losses are generated during the turn-on and the turn-off states, therefore the losses depending on the turn-on and turn-off energy, and the switching frequency [4] [9].

$$P_{loss\_switch} = \left(E_{on} + E_{off}\right) f_{sw}. \tag{3}$$

where

E_{on} is the turn on energy of the semiconductor;

E_{off} is the turn off energy of the semiconductor;

f_{sw} is the switching frequency.

In case of diode (body diode) is distinguishable (in generally) the conduction and the switching loss (4), (5) [9]:

$$P_{loss\_cond} = V_d I_f, \tag{4}$$

where

V_d is the diode drop voltage;

I_f is the current through the diode.

$$P_{loss\_switch} = \frac{1}{4} Q_{rr} V_r f_{sw}, \tag{5}$$

where

Q_{rr} is the reverse recovery charge;

V_r is the reverse drop voltage;

f_{sw} is the switching frequency.

4. Parallel connection of the power semiconductors

The high-power systems make the parallel connection of semiconductors necessary. This is possible in case of discrete or modular design, too. The aim of parallel connections is increasing of the system current or increasing the redundancy of the system. The current shares equal between the semiconductors. If the current is equally, then the warming and the losses will be equal, too. If that does not happen, then thermal difference is generated, which can increase the losses or damage the semiconductor. It is very important, that the semiconductors should have to have the same value gate of resistor.

However, the parallel connection can decrease the equivalent resistance, on the same time it means more capacitance, and hence, more charge. This will increase the switching losses at high switching frequency.

4.1. Parallel connection of MOSFET's

In case of MOSFET the drain-source channel can be represented by a resistor, which behaves as PTC resistor. If the temperature increases on the semiconductor, then R_{DS} will be higher. Thermal differential is not formed, because the higher resistivity MOSFET conducts smaller current and forcing the higher current to the other semiconductor. This property makes the parallel connection of the MOSFET easy. In case of GaN-based semiconductor it is the same way (self-balancing) [10].

4.2. Parallel connection of IGBT's

In case of parallel connection of IGBT there are two important properties: the saturation voltage and the U_{GE} and I_D transfer characteristic. The latest is the same, as the MOSFET $R_{DS(on)}$ parameter. Both parameters are depending on the temperature. The thermal coefficient IGBT is a little bit more complex, as in case the MOSFET. The transfer characteristic of IGBTs contains an isothermal point: above this point the IGBT has PTC behavior (in case of higher temperature the on-state resistance is increasing), under this point the IGBT has NTC behavior. During the design the aim is the equal distribution of the load, therefore it must strive the PTC behavior [11].

The behavior of the semiconductor depends on the U_{GE} voltage. If the setting of U_{GE} voltage is right, then the semiconductor is self-regulated: if any IGBT has higher current, then the temperature on the IGBT will be higher, therefore the on-state resistance is increasing. If the aim of symmetrical operation, then each IGBTs must have the same gate resistor [11].

5. Conclusion

In this paper the current semiconductor trend in the electric drive system has been introduced and the opportunities of the future has been investigated. Nowadays the industry and the developers use IGBTs in high power systems, but the wide bandgap semiconductors offer new opportunities. The range, the efficiency and the switching frequency are increasing in the electric drive systems, therefore the size of the circuits and reactant parts, and their heat sink will be smaller, which mean smaller volume, size and weight. These new capabilities enable the designers to find an optimal approach to enhance the performance, reduce the losses. *Fig. 4.* illustrates a possible future trend.

In the future will be designed a step-down converter with GaN E-HEMT transistor and Linear-quadratic controller for the best efficiency.

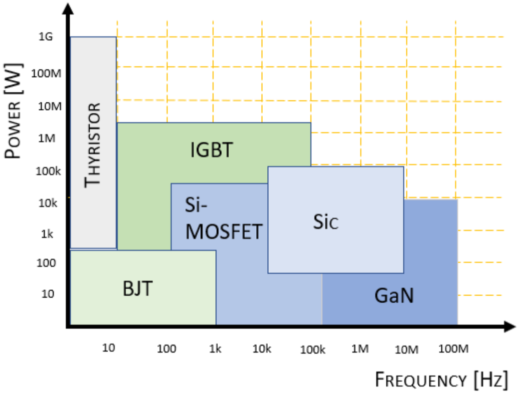


Figure 4. Possible future trend in semiconductors [12]

Acknowledgement

The publishing of this paper was supported by the project “EFOP-3.6.1-16-2016-00017 Internationalisation, initiatives to establish a new source of researchers and graduates and development of knowledge and technological transfer as instruments of intelligent specialisations at Széchenyi István University”.

References

[1] In a Tesla Model S, there is no IGBT packing trick
URL https://www.pntpower.com/on-tesla-electric-vehicles-semiconductor-packaging/

[2] E. Barbarini, STMicroelectronics SiC Module Telsa Model 3 Inverter, Power Semiconductor report. 2018.
URL https://www.systemplus.fr/wp-content/uploads/2018/06/SP18413-STM_SiC_Module_Tesla_Model_3_Inverter_sample-3.pdf

[3] SiC Power Devices and Modules, ROHM Semiconductor. Application note. 2013.
URL https://www.rohm.com/documents/11308/2420552/SiC_power_device-catalog.pdf

[4] A. S. Abdelrahman, Z. Erdem, Y. Attia, M. Z. Youssef, Wide Bandgap Devices in Electric Vehicle Converters: A Performance Survey, Canadian Journal of Electrical and Computer Engineering, 41 (1) (2018) pp. 45-54.
doi: 10.1109/CJECE.2018.2807780

[5] L. Stephen, A. Robert, Fundamentals of Gallium Nitride Power Transistors. Application note. 2011.
URL https://epc-co.com/epc/Portals/0/epc/documents/product-training/Appnote_GaNfundamentals.pdf

[6] Design with GaN Enhancement mode HEMT. GN001 Application note. 2018.
URL https://gansystems.com/wp-content/uploads/2018/02/GN001_Design_with_GaN_EHEMT_180228-1.pdf

[7] Z. John Shen, Power Semiconductor Devices for Hybrid, Electric, and Fuel Cell Vehicles, Proceedings of the IEEE, 95(4) (2017) pp. 778 – 789.
doi: 10.1109/JPROC.2006.890118

[8] M.H. Rashid, Power Electronics Handbook, 3rd Edition, Butterworth-Heinemann, 2007.

[9] H. Kakitani, R. Takeda, Selecting the Best Power Device for Power Electronics Circuit Design through Gate Charge Characterisation, Elektronika, 56 (10) (2015) pp. 55-59.

[10] Z. Puklus, Power electronics, Universitas-Győr Nonprofit Kft, 2007, in Hungarian

[11] Paralleling of IGBTs, ON Semiconductor. Application note. 2014. URL https://www.onsemi.com/pub/Collateral/AND9100-D.PDF

[12] A. Bhalla, Practical considerations when comparing SiC and GaN in power applications. Application note. 2018. URL http://unitedsic.com/wp-content/uploads/2018/07/Practical-considerations-when-comparing-SiC-and-GaN-in-power-applications.pdf

Acta
Technica
Jaurinensis

Vol. 12, No. 1, pp. 26-54, 2019
DOI: 10.14513/actatechjaur.v12.n1.489
Available online at acta.sze.hu

Railroad Ballast Particle Breakage with Unique Laboratory Test Method

E. Juhász[1], Sz. Fischer[1]

[1]Department of Transport Infrastructure and Water Resources, Faculty of Architecture, Civil Engineering and Transport Sciences, Széchenyi István University
Egyetem tér 1., H-9026 Győr, Hungary
e-mail: juhasz.erika@sze.hu, fischersz@sze.hu

Abstract: This paper demonstrates the results in the research topic of the railway ballast particles' breakage test with unique laboratory test. The most railway lines in the world have so called traditional superstructure (ballasted tracks). In the past few years there were a lot of railway rehabilitation projects in Hungary, as well as abroad. Nowadays cannot be expected that there is enough quantity of railway ballast in adequate quality, because of the modifications and restrictions in the related regulations in Hungary since 2010. In Hungary there are only a few quarries which are able to ensure adequate railway ballast material for construction and maintenance projects for speed values between 120 and 160 km/h. This may cause supply and quality risk in production of railway ballast. The authors' research's main goal is to be able to simulate the stress-strain effect of ballast particles in real and objective way in laboratory conditions as well as in discrete element modelling.

Keywords: ballast, breakage, laboratory test, railway, degradation

1. Introduction

An article was published [1] in 2015 with results of an R&D on individual breakage test method in laboratory related to railway ballast. Since that several other publication was published in this topic [2-6]. The authors would like to supplement that documents with actual, up-to date outcomes.

The rock physical suitability of railway ballast materials is determined by laboratory tests in in the EU, formulated in the same product standard.

There are two types of standardized tests in the aspect of rock physic characteristics of railway ballast:

- Micro-Deval abrasion test according to MSZ EN 1097-1:2012 [7],
- Los Angeles abrasion test in accordance with MSZ EN 1097-2:2010 [8].

These are determined in the MSZ EN 13450:2003 product standard [9].

These two test types are absolutely suitable for satisfy defining the abrasion characteristics of a given aggregate sample and for ensuring the production stability in the quarries and these are indispensable to guarantee the required quality and to ensure the checking of the quality level in case of ready constructed railway tracks. However, it's not suitable for modelling the railway loads (i.e. loads from vehicles and other effects) in a real way.

The authors worked out an individual laboratory test method [1, 3], because other test methods can't consider the abrasion and breakage (real particle degradation) due to dynamic force and vibration.

The results of the unique laboratory tests with the required limits in standards are compared with the related regulation of MÁV (Hungarian Railways) [10]. Required time intervals of ballast screening are able to be calculated according to laboratory test results.

It is known that No. MÁV 102345/1995. PHMSZ in accordance with the decree [10], the "Constructions for superstructure facilities and quality standards for bedding instruction" has been tightened on the basis of the 4th amendement [10], which came into force in January 2010. According to the 3rd amendement of 2008, there was a (positive) tolerance range for the Los Angeles breakdown and the Micro-Deval wear, which was deleted in 4th amendement. The values for the speed categories are also partially have changed, usually tightened (*see Table 1. and Table 2.*).

Table 1. Requirements to LA_{RB} values [10]

strength requirement	LA_{RB}			
	2008 - 2009		from 2010	
allowed speed (km/h)	requirement	allowed difference	requirement	allowed difference
V > 160	16	+2 (negative is not limited)	16	-
160 ≥ V > 120	16	+4 (negative is not limited)	16	-
120 ≥ V ≥ 80	16	+4 (negative is not limited)	16	-
80 > V ≥ 40	24	+4 (negative is not limited)	20	-
V < 40	24	+4 (negative is not limited)	24	-

Table 2. Requirements to $M_{DE}RB$ values [10]

strength requirement	$M_{DE}RB$			
	2008 - 2009		from 2010	
allowed speed (km/h)	requirement	allowed difference	requirement	allowed difference
V > 160	11	+2 (negative is not limited)	11	-
160 ≥ V > 120	11	+4 (negative is not limited)	11	-
120 ≥ V ≥ 80	11	+4 (negative is not limited)	15	-
80 > V ≥ 40	15	+4 (negative is not limited)	15	-
V < 40	15	+4 (negative is not limited)	15	-

2. History of the research

The research topic has prestigious international literature and sources. Foreign researchers dealt with different areas and they worked out different methods as follows:

- laboratory tests [11, 12, 13, 14, 15, 16, 17, 18, 19]
- finite element modelling (FEM) [13],
- discrete element modelling (DEM) and/or 3D particle generation [19, 20],
- in-situ tests in railway tracks [21].

The researchers worked out several special parameters, constants and indexes that helped the progression of the research (e.g. Marshal, Hardin, Lee and Farhoomand breakages, BBI index, B_R index, etc.).

An international literature review was carried out by the authors and from the results the following main themes and methods were taken into consideration:

- searching of relationship between cohesion as well as inner friction anger, railway ballast aggregate abrasion, water permeability of material and its layer [16];
- definition of relationship between Particle Size Distribution (PSD) and particle degradation phenomenon of ballast aggregate, as well as definition of better PSD for real loadings [15, 16];
- research of ‘angularity breakage’ phenomena [18];
- DEM models were validated and DEM generations method of much more realistic particle shapes was investigated [22];
- measurement of railway ballast’s breakage, implementation of laboratory and field tests with and without geosynthetic inclusions [13, 14, 15, 21];
- research of ballast particle breakage due to tie tamping [11, 12];
- investigation of ballast with glued technique [17].

The authors collected several significant results at the international literature review which can be read in the previous publication [4].

3. The laboratory test's procedure and parameters

The base of the procedure is a special laboratory dynamic actuator The laboratory test method was developed as a part of an R&D financed by Colas Északkő Ltd. and a report more publications – that were written in this topic [1, 2, 3, 4, 5, 6].

In 2017 and 2018 the testing and evaluating method was accomplished by specify more precise deterioration process, considering only determined particle fraction, etc. The authors used the following parameters during measurements and evaluation:

- two different types of railway ballast samples from andesite material and from different quarries
- the specimens are in accordance with MSZ EN 13450:2003 [9], A type, 31,5/50 mm, the authors received from Colas Északkő Ltd.
- the specimens have the following stone physic parameters (laboratory test were done by accredited laboratory of Colas Északkő Ltd.):
  - sample No. 1: $LA_{RB} = 19\%$, $M_{DE}RB = 17\%$;
  - sample No. 2: $LA_{RB} = 16\%$, $M_{DE}RB = 4\%$;
- dynamic tests with pulsator in different cycles (i.e. until 0.1, 0.2, 0.5, 1.0, 1.5, 3.0 and 5.0 million cycles), in every test with only fresh ballast material with particle fraction $d \geq 22.4$ mm (before pulsating $d < 22.4$ mm particles were screened out and they were not put back), where *d* is the size of the particle; in the 2014 series of measurements, the particle sizes below 22.4 mm were left in the tests, this was the reason why it is not possible to compare the current measurement results with the old ones;
- determination of PSD curves with screening related to sub-samples Before Pulsating (BP) test;
- determination of PSD curves with screening related to sub-samples After Pulsating (AP) test.

3.1. Presentation of the individual fatigue laboratory test

The individual laboratory testing method is a dynamic pulsating test for that the six lower frames of a 10-level steel shear box were used [1]. Frames were fixes together with steel metric screws. They prevent the horizontal relative displacements. The shear box including some steel rolls which were not fixed to the down side of the bottom frame.

The built-up layer structure is the following *(see Table 3)*.

Table 3. The built-up layer structure

steel loading plate
D=300 mm steel plate with circular
shape (D=diameter)
0.46 x 0.42 m

<table>
<tr><td>*30-cm-thick-layer*
wooden sleepers
around the crushed stone samples</td><td>*simple layer Viacon PP TC 1200* ***geotex.***</td><td>*30-cm-thick-layer*
crushed stone
cross section 0.46 x 0.46 m</td><td>*simple layer Viacon PP TC 1200* ***geotex.***</td><td>*30-cm-thick-layer*
wooden sleepers
around the crushed stone samples</td></tr>
<tr><td colspan="5">*simple layer*
heat treated, non-woven, high strength geotextile with 1200 g/m^2 mass
type: Viacon GEO PP TC 1200
on the whole 1.0 x 1.0 m area</td></tr>
<tr><td colspan="5">*10-cm-thick-layer*
sand
type: E_2, 20.42 MPa according to MSZ 2509-3:1989 [23]
on the whole 1.0 x 1.0 m area</td></tr>
<tr><td colspan="5">*simple layer*
150 g/m^2 mass geotextilie
type: Naue Secutex 151 GRK
on the whole 1.0 x 1.0 m area</td></tr>
<tr><td colspan="5">*20-cm-thick layer*
eXtruded Polystirol (XPS)
type: Austrotherm Thermoplan
sheets on the whole 1.0 x 1.0 m area</td></tr>
</table>

The samples (railway ballast) were put in the middle of the shear box into the 0.46×0.46 m area and 0.30 m high space where wooden sleepers are around. The structure can be seen in *Table 1*. Reducing and excluding wall effect the inner sides of wooden sleepers were covered with 1200 g/m^2 mass geotextile layers (where stone and wooden sleepers would interact). A loading plate from steel (size: 46×42 cm and D=300 mm circular steel plate) were put onto ballast samples to be able to achieve uniform load distribution.

The assembly without loading plates can be seen in *Figures 1-5*.

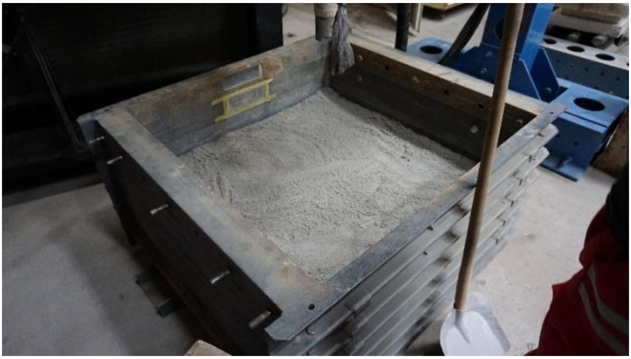


Figure 1. The 10-cm-thick-layer sand on the XPS layer

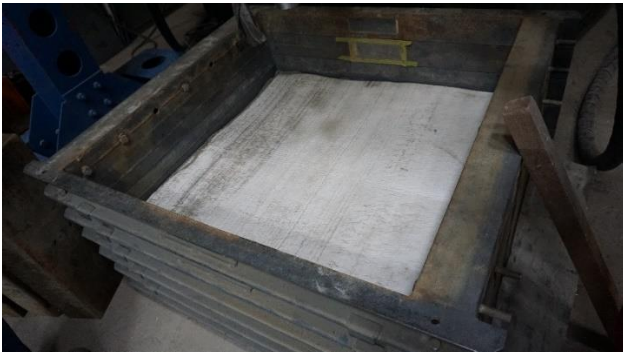


Figure 2. The high strength geotextilie on the sand layer

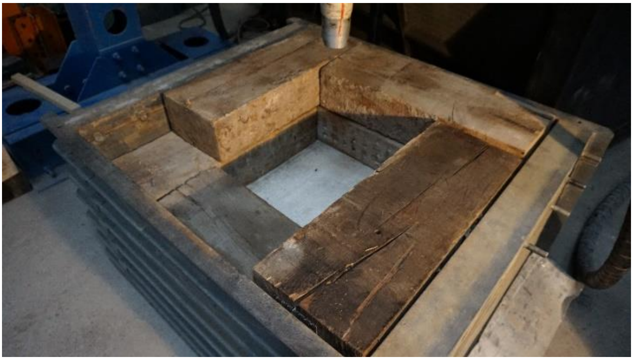


Figure 3. The wooden sleepers in the shear box

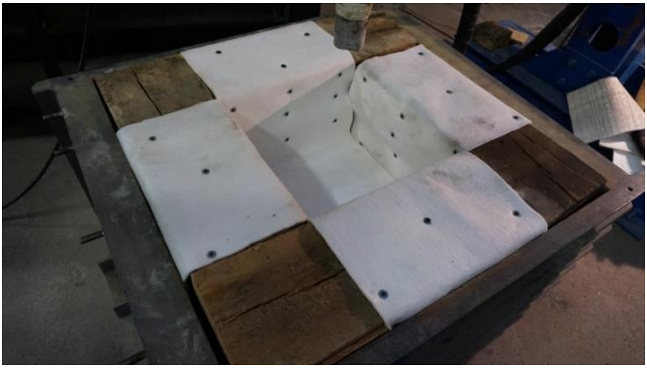


Figure 4. The geotextilie among the wooden sleepers

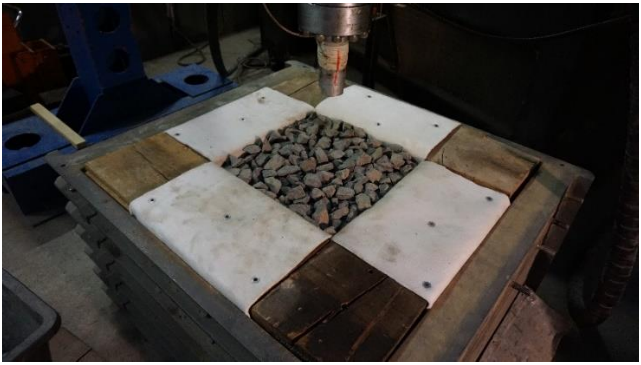


Figure 5. The sample of ballast material in place

Laboratory measurements were executed with dynamic actuator in previously defined loading cycles. Laboratory test parameters (frequency, load values, etc.) were published in [1, 3] papers in detailed formats, they are not written here due to limited space. After pulsating tests PSD (particle size distribution) data sets were determined (measured) beside them several parameters (see below) were calculated [1, 2, 3]:

- F_V (%) (see eqs. (1-5));
- *BBI* (see eq. (6));
- B_R, (B_R is a parameter is similar to BBI, but it considers different areas in PSD [24]);
- *d<22.4 mm* in mass percentage;

- *$d<0.5$ mm* in mass percentage;
- *$d<0.063$ mm* in mass percentage;
- *d_{60}/d_{10} ratio* (where d_{60} is the particle size related to 60% in PSD curve, d_{10} is the particle size related to 10% in PSD curve, and *d_{60}/d_{10} ratio* means the ratio of $(d_{60}/d_{10})_{AP}/(d_{60}/d_{10})_{BP}$;
- *C_C ratio* (where $C_C=d_{30}^2/(d_{60}\times d_{10})$, d_{30} is the particle size related to 30% in PSD curve, and *C_C ratio* means the ratio of $(C_C)_{AP}/(C_C)_{BP}$;
- *M ratio* (where M is a special shape factor of PSD curve of railway ballast [25], and *M ratio* means the ratio of M_{AP}/M_{BP};
- *λ ratio* (where λ is a special shape factor of PSD curve of railway ballast that considers standard ballast PSD, as well [25], and *λ ratio* means the ratio of $\lambda_{AP}/\lambda_{BP}$).

Calculations of F_V and BBI parameters have to be explained, eqs. (1-6) give the meaning of these parameters [1, 2, 3, 14, 15, 26].

$$F_V = 0.4 \cdot F_{19} + 0.3 \cdot F_{6.7} + 0.2 \cdot F_{1.18} + 0.4 \cdot F_{0.15}, \quad (1)$$

$$F_{19} = \frac{D_{19}}{100} \cdot 27, \quad (2)$$

$$F_{6.7} = \frac{D_{6.7}}{100} \cdot 18, \quad (3)$$

$$F_{1.18} = \frac{D_{1.18}}{100} \cdot 11.5, \quad (4)$$

$$F_{0.15} = \frac{D_{0.15}}{100} \cdot 5.5, \quad (5)$$

where “D” is the fallen mass percentage through the given diameter sieve.

$$\mathrm{BBI} = \frac{A}{A+B}, \quad (6)$$

where A is the area between the initial and final PSD curves [14, 15], B is the area between the arbitrary boundary of maximum breakage line and final PSD curve [14, 15].

The authors computed the required time intervals of ballast screening with the help of deterioration process obtained from parameters above. The prescribed values from standards were also computed.

4. Recent results

The two ballast samples tested until 5-5 million cycles with dynamic pulsating laboratory test in more phases. The authors calculated the necessary parameters from PSD data sets are plotted in diagrams as a function of pulsating cycles in *Figure 6-25*.

Following diagrams show the parameters in the consideration of the maximum 5-5 million loading cycles, as well as the results in *Figure 6-25*.

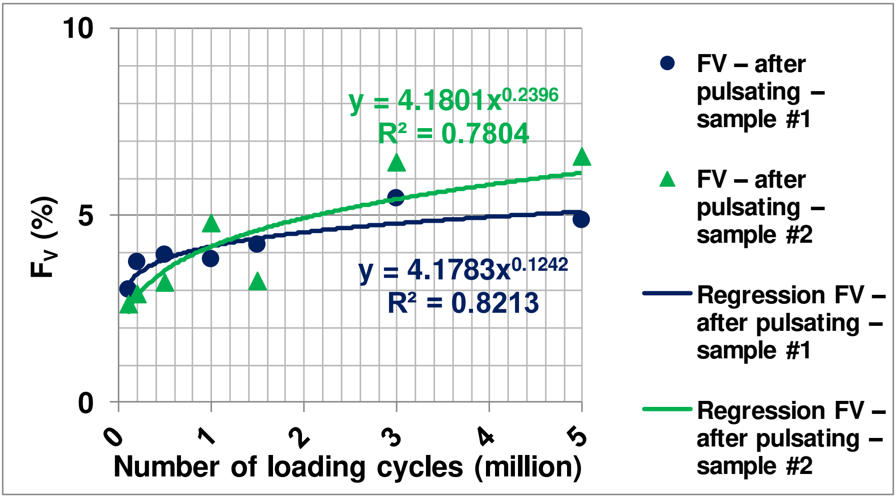


Figure 6. Results of the individual laboratory test – F_V (%) as a function of number of loading cycles; with power regression functions

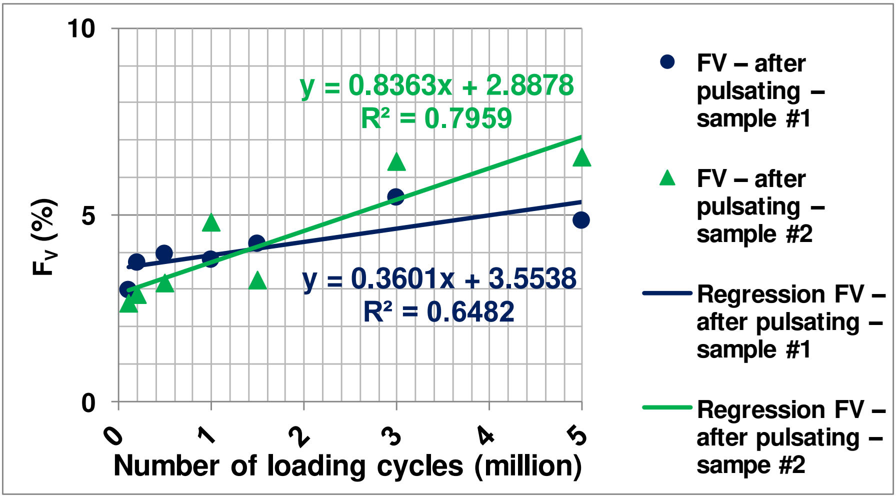


Figure 7. Results of the individual laboratory test – F_V (%) as a function of number of loading cycles; with linear regression functions

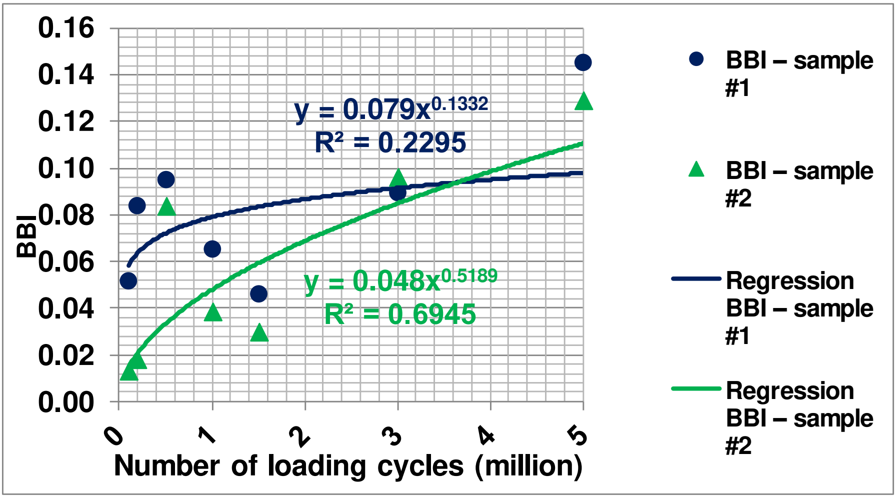


Figure 8. Results of the individual laboratory test – BBI as a function of number of loading cycles; with power regression functions

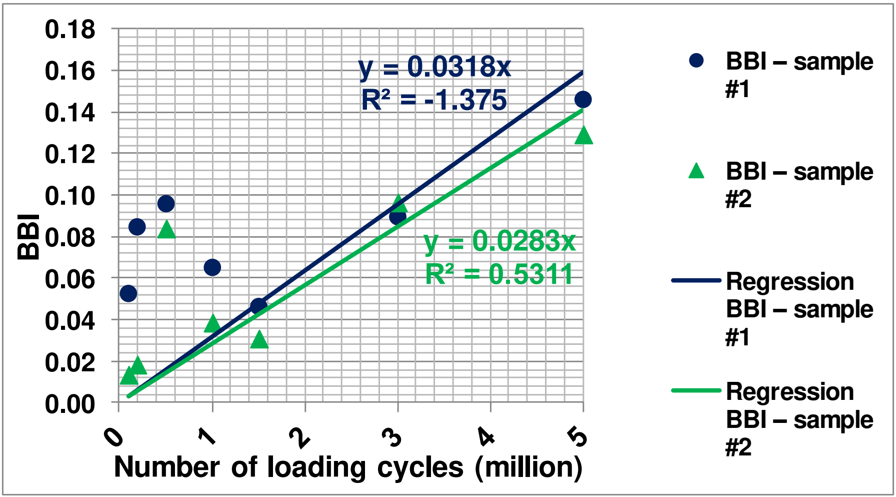


Figure 9. Results of the individual laboratory test – BBI as a function of number of loading cycles; with linear regression functions

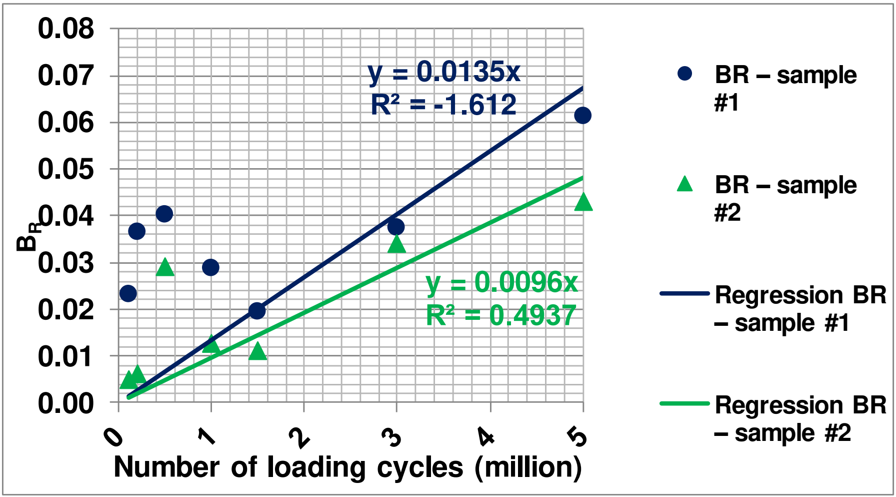


Figure 10. Results of the individual laboratory test – B_R as a function of number of loading cycles; with power regression functions

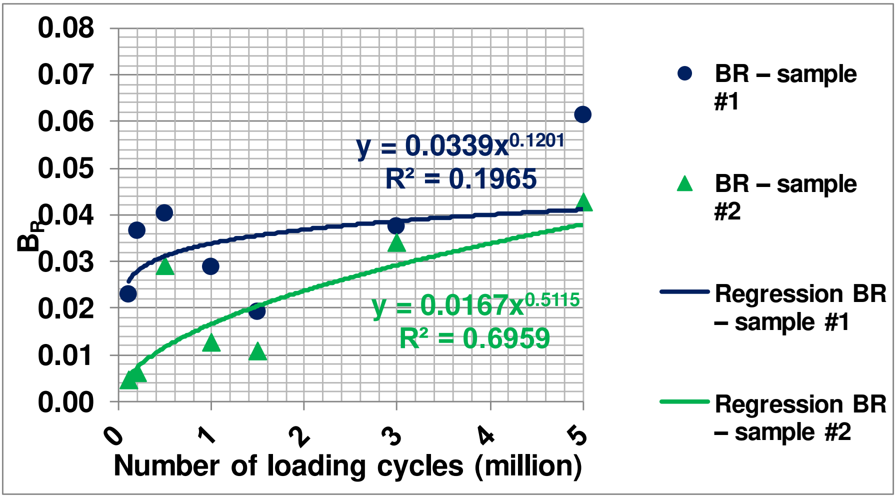


Figure 11. Results of the individual laboratory test – B_R as a function of number of loading cycles; with linear regression functions

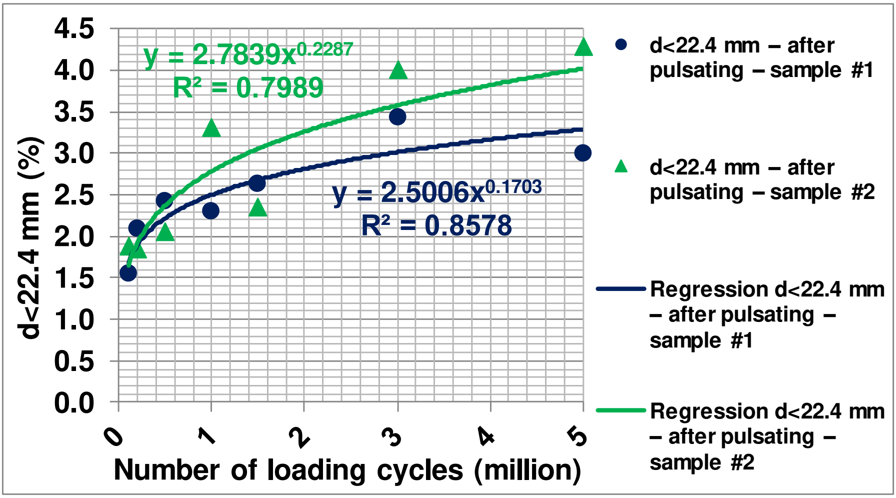


Figure 12. Results of the individual laboratory test – d<22.4 mm (%) as a function of number of loading cycles; with power regression functions

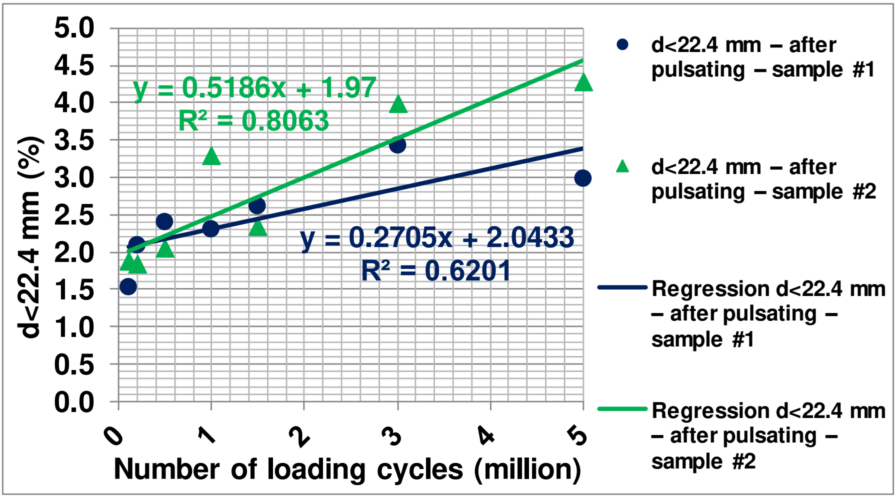


Figure 13. Results of the individual laboratory test – d<22.4 mm (%) as a function of number of loading cycles; with linear regression functions

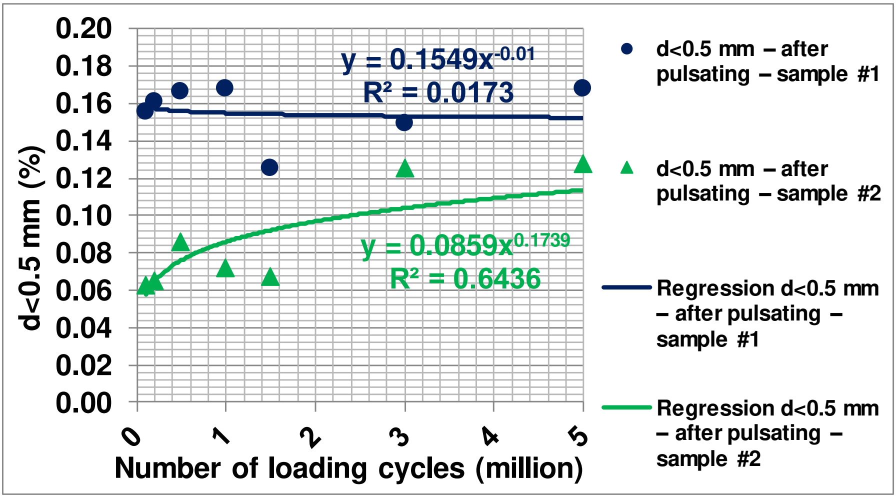


Figure 14. Results of the individual laboratory test – d<0.5 mm (%) as a function of number of loading cycles; with power regression functions

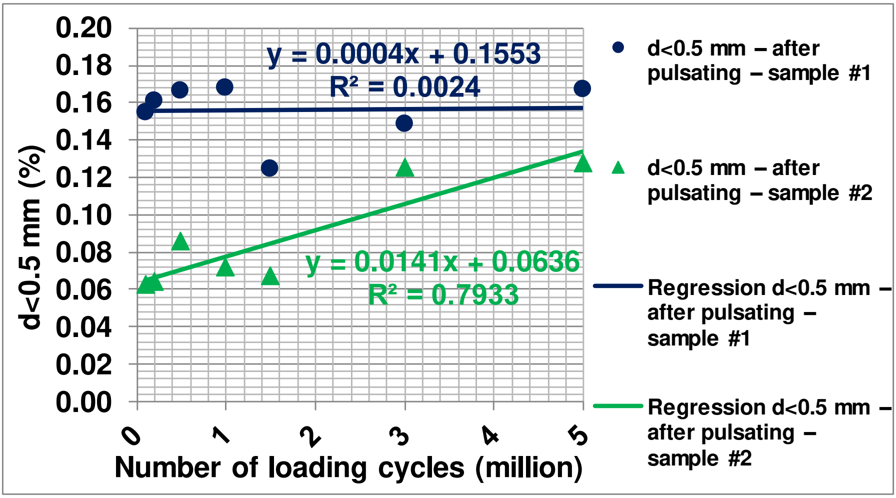


Figure 15. Results of the individual laboratory test – d<0.5 mm (%) as a function of number of loading cycles; with linear regression functions

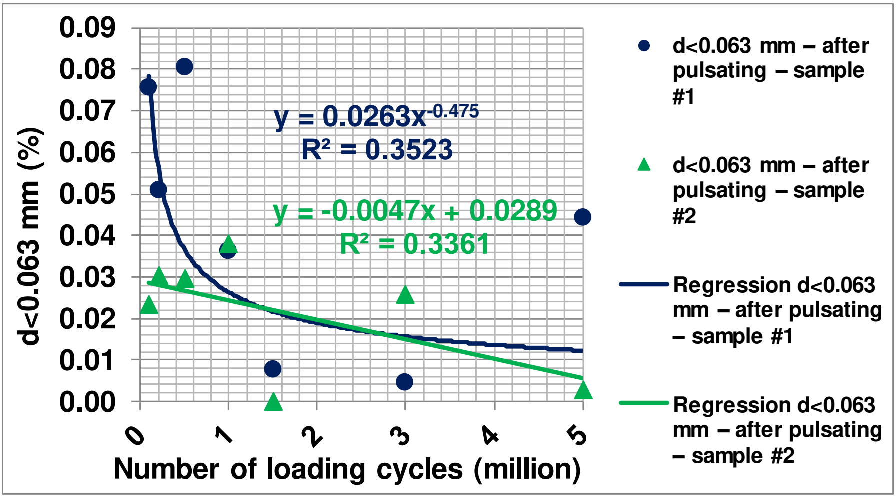


Figure 16. Results of the individual laboratory test – d<0.063 mm (%) as a function of number of loading cycles; with power and linear regression functions

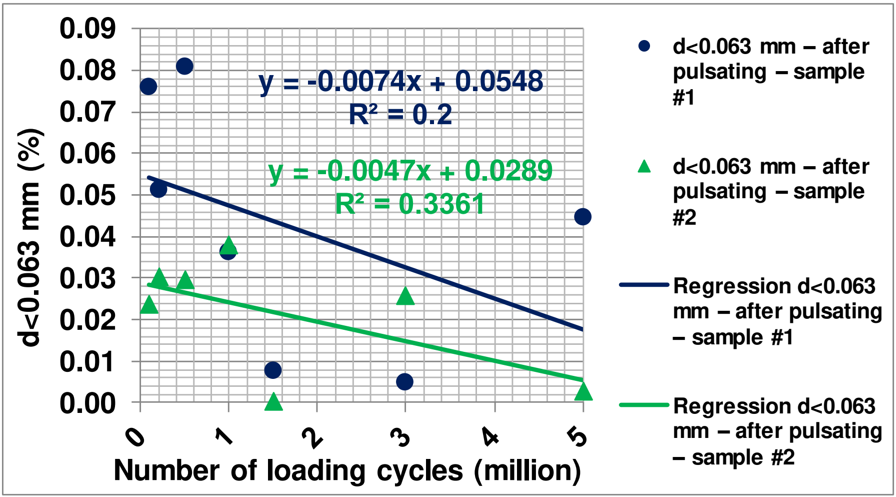


Figure 17. Results of the individual laboratory test – d<0.063 mm (%) as a function of number of loading cycles; with linear regression functions

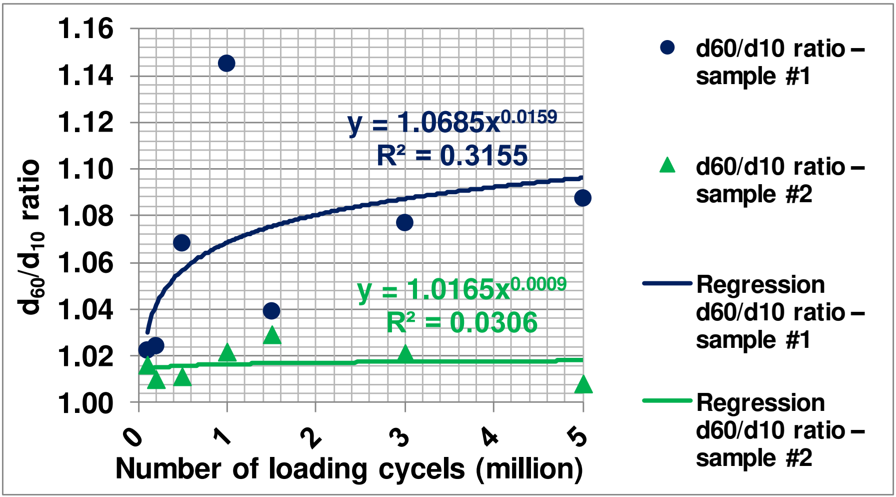


Figure 18. Results of the individual laboratory test – d_{60}/d_{10} ratio as a function of number of loading cycles; with power regression functions

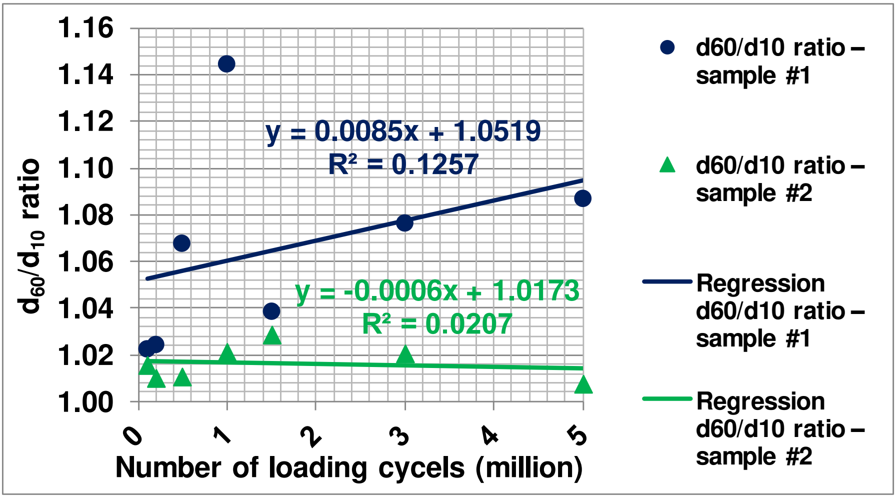


Figure 19. Results of the individual laboratory test – d_{60}/d_{10} ratio as a function of number of loading cycles; with linear regression functions

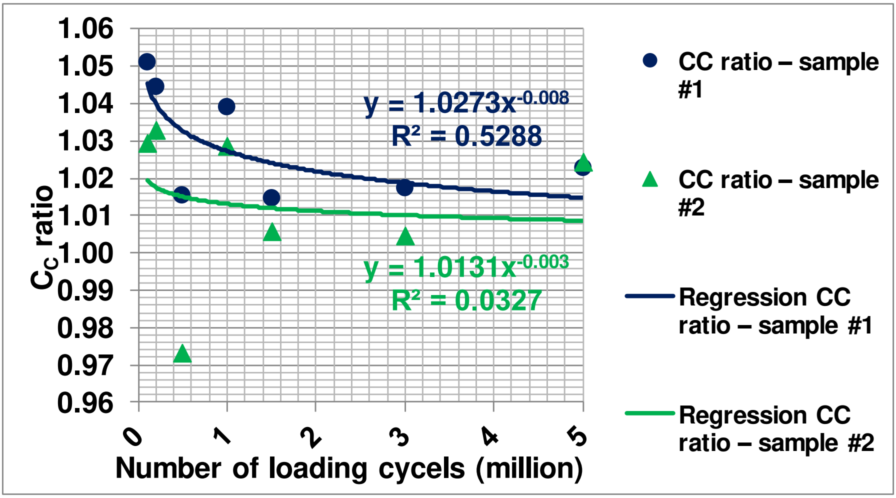


Figure 20. Results of the individual laboratory test – C_C ratio as a function of number of loading cycles; with power regression functions

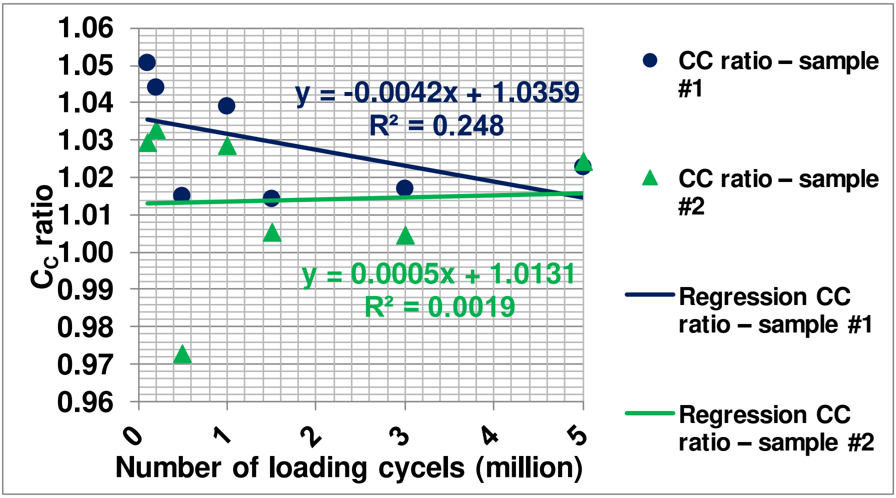


Figure 21. Results of the individual laboratory test – C_C ratio as a function of number of loading cycles; with linear regression functions

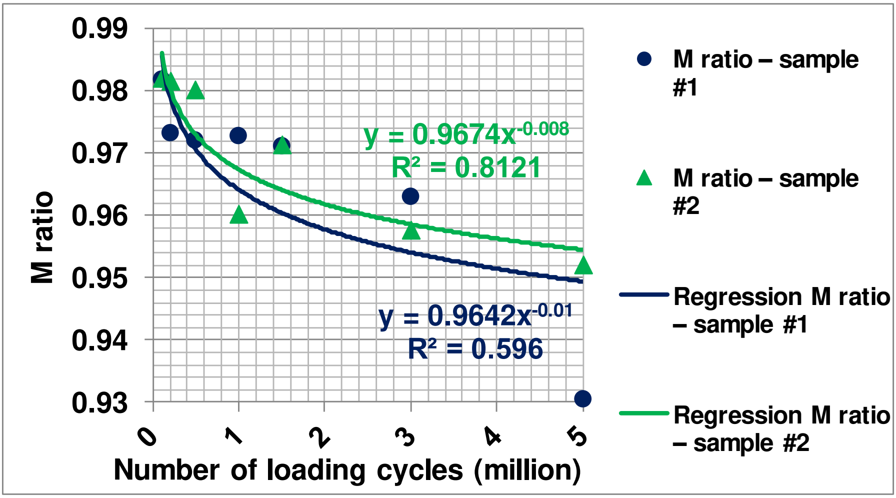


Figure 22. Results of the individual laboratory test – M ratio as a function of number of loading cycles; with power regression functions

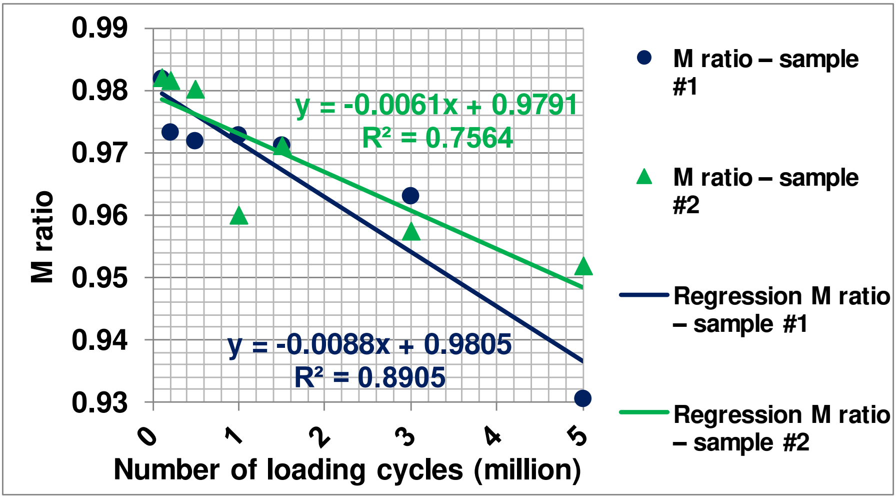


Figure 23. Results of the individual laboratory test – M ratio as a function of number of loading cycles; with linear regression functions

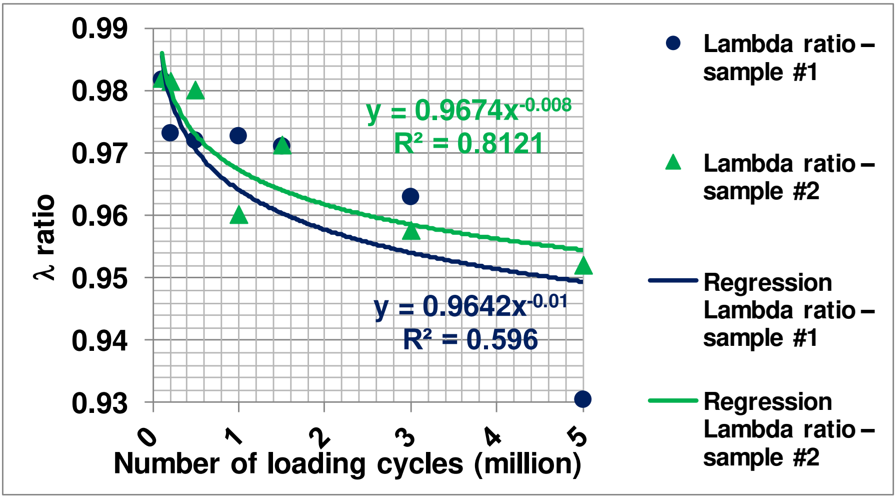


Figure 24. Results of the individual laboratory test – λ ratio as a function of number of loading cycles; with power regression functions

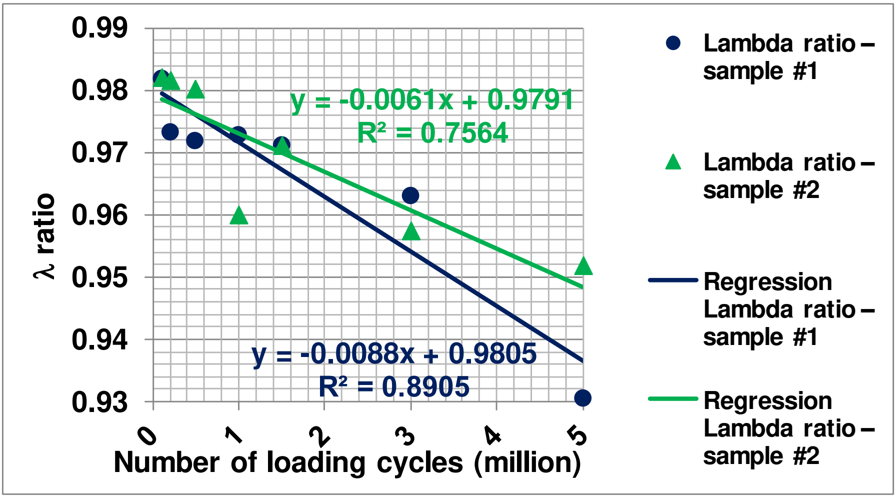


Figure 25. Results of the individual laboratory test – λ ratio as a function of number of loading cycles; with linear regression functions

There are significant correlations in four calculated parameters from ten cases (linear and power regression functions, independent variable are the number of loading cycles):

- F_V: the parameter recommended by the South African Railways [26]. This indicates the necessity of ballast screening. It can be approximated by power regression function of the number of loading cycles (see *Figure 6*). When it reaches the 5 million loading cycles, the F_V equals are 5.0 and 6.5%. It is an interesting fact that better ballast sample deteriorates faster than the less good ballast sample. This means the considering stone physics attribution sample #2 has lower LA_{RB} and $M_{DE}RB$ values then sample #1. Ballast screening has to be executed if F_V=80% according to literature [26]. In case power regression functions are considered and number of loading cycles is calculated related to F_V=80%, they are 2.1×10^{16} cycles and 2.24×10^{11} cycles (for sample #1 and sample #2, respectively). These values means 8-10 million loading cycles (through-rolled axles) what are unrealistic high. On Kelenföld-Hegyeshalom state border, No. 1 railway mainline in Hungary the loading is approximately 15 million MGT/year, that is 666.667 axles/year – and it taking into consideration the common used time interval of ballast screening (12-15 years) [26]. In this aspect, results obtained from analysis of F_V parameter and its extrapolation cannot be used dependably.

- *BBI*: the parameter introduced by Indaratna and Lackenby [14, 15] to be able to calculate and assess the changing of PSD of tested samples and their quality. For calculation the BBI index, the authors have to determine the BP and AP values and with these the PSD curves. There is linear regression correlation between BBI and number of loading cycles (see *Figure 9*), but it has to be mentioned that in case of sample #1 the R2 coefficient is negative. The negative value mathematical means that the given correlation is worse than the *y* has constant function. It has to be mentioned that BBI(0)=0 was a border condition. If measured data before 1.0 million loading cycle are analysed they are outliers from the defined linear trend. The measurements are needed to repeat, because of the ability to evaluate fair. BBI values are approximately 0.15% and 0.20% at 5 million loading cycles related to tested ballast material samples. It is interesting too that sample #2 has higher tangent (steeper slope) than sample #1 (see F_V, too). The authors neglect the negative R_2 value, because after 1.5 million loading cycles the given linear trend can be assumed). According to the literature [14, 15], if the BBI=1.0, ballast screening is needed. In case number of loading cycles related to both samples are computed in the accordance BBI=1.0, the results are 31.45 and 25.84 million cycles (for sample #1 and #2, respectively). That means approximately 47-year and 39-year time intervals.
- *B_R*: this is a similar parameter with the BBI parameter, because reference area is calculated like in case of BBI [24]. In this way the linear regression functions and correlation are very close to BBI's results. There is no technical recommendation for ballast screening related to B_R (see *Figures 10-11*).
- *d<22,4 mm is mass percentage*: between the *d* parameter and the number of loading cycles there are power regression relationships (*see Figures 12-13*). This parameter's values are approximately 3.0% and 4.25%, if the number of loading cycles is 5×10^6 related to investigate andesite materials. This is mentioned above, but it is a surprising fact again that the sample #2 with better properties has quicker deterioration than sample #1. According to literature [26] ballast screening has to be done if *d*<22.4 mm is equal or higher than 30%. In case numbers of loading cycles has to be determined they are the following: 2.17×10^{12} and 3.27×10^{10} cycles (for sample #1 and #2, respectively) that are unrealistic values.
- *d<0,5 mm is mass percentage*: between the *d* parameter and the number of loading cycles there is no power and linear regression relationship (*see Figures 14-15*).

- *$d<0.063$ mm is mass percentage*: between the *d* parameter and the number of loading cycles there is no power and linear regression relationship (*see Figures 16-17*).
- *d_{60}/d_{10} ratio parameter:* between the *d_{60}/d_{10} ratio parameter* and the number of loading cycles there is no power and linear regression relationship (*see Figures 18-19*).
- *C_C ratio parameter:* between the *C_C ratio parameter* and the number of loading cycles there is no power and linear regression relationship (*see Figures 20-21*).
- *M* and *λ ratio parameters*: these parameters can be specified by linear regression function (*see Figures 22-25*). The tangent of regression function is higher in case of sample #2 than sample #1 (the sign is positive for sample #2 and negative for sample #1). The authors analysed the data more detailed and determined that there is an outlier point in case of sample #2 (at 5 million cycles, both *M* and *λ* ratio parameters), the reason of it has to be searched by additional tests.

5. Summary, outlook, future scope

There are significant correlations in case of just four from the nine calculates parameters if the independent variable is the number of loading cycles (linear or power). With the other five parameters we did not achieve any results.

The authors considered the following derelictions related to calculation of time intervals between ballast screenings:

- in the whole ballast cross section comparable amount of breakage is not formulated as the one that was measured in referred laboratory tests (e.g. there is hardly no breakage in the ballast shoulder, etc.);
- machine-made and/or manual tamping occurred breakage;
- only 225 kN axle load was taken into consideration (it is true for freight trains, for passenger trains about 180 kN value would be more realistic);
- other ballast polluting effects (e.g. dust, concrete sleeper abrasion, breakage, in case of water pockets the increase of fine particle content in the ballast bed because of evolving pumping effect due to repeated dynamic load, etc.);
- deterioration effect accelerated by substructure or superstructure defect;
- effects of other dynamic loadings (e.g. welts, rail joints, turnout frogs [27, 28, 29, 30, 31, 32]);
- effects of track geometry and its degradation [33].

This paper introduced the research problem related to railway ballast particle degradation. The publication sentenced that the individual laboratory testing method can be suitable for measure and evaluate ballast materials' breakage using dynamic pulsator. This procedure can give opportunity that ballast samples are tested in more realistic circumstances than during standardized abrasion tests.

The authors summarized the up-to-date results of exhausting international literature in this research topic; they made a remarkable successes related to ballast particle breakage.

It can be stated that better ballast material (in the aspect of stone physics) does not deteriorate slower than the worse one. Parameters that are used internationally were calculated, as well as the deterioration process was approximated linear or power regression functions related to all parameters.

The authors defined the time interval values of ballast screening based on technical prescriptions, standards and handbooks. This calculation could be done for F_V, *BBI* and d<22.4 parameters. Only *BBI* gave nearly acceptable results, in case of the other parameters the results are not realistic, they can't be accepted.

In some cases, additional control measurements have to be accomplished in the laboratory to be able to assess the measured data.

The authors would like to search the correlation (relationship) – as future scope – between standardized parameters (Los Angeles and Micro-Deval abrasions), the prognosticated time interval between ballast screening, as well as the results from their laboratory tests. In the beginning of 2019 a modified layer structure will be considered because the extruded polystyrol sheets were significantly deformed during the dynamic tests. A stiffer and harder layer (e.g. granular protection layer or steel/concrete plate, maybe) can be modified laboratory tests' results in better way, as well as difference between substructure circumstances with XPS sheets and stiffer layer is able to be published.

The time requirement of newly developed testing method is significantly high, in this way the authors would like to execute laboratory tests with lower time demand (e.g. particle splitting tests), so relevant statements can be sentenced sooner. The authors would like to combine this splitting test with a full-field 3D shape measurement (ATOS fringe projection system) and/or X-ray measurement technique. The measurement method need to be work out [18, 34, 35, 36, 37, 38, 39, 40, 41, 42].

Beside them field tests are planned in the Hungarian railway lines. The authors would like to collect samples from old railway lines where ballast aggregates have known PSD at time of construction. The actual PSD can be definable and the changing can be determined, too. In case a lot of these kinds of measurements are able to be performed the comparison (not only field samples but the others from laboratory dynamic pulsating tests) can supply valuable results.

In the laboratory measured particle breakage values are much higher than the values in real circumstances in tracks, either in tracks with maintenance (ballast screening) demand. The reason is the only one type of loading form used in laboratory. The authors would like to develop their methodology to be able to assess the particle degradation more realistic. Tamping machines also break ballast particles, so this kind of effect is needed to be considered in the future research. Other additional dynamic loading effect can't be neglected in sophisticated methods, e.g. surroundings of rail welts, rail joints, as well as switch frogs where higher ballast breakage should be expected [27, 28, 29, 30, 31, 32]. Rubber coated and bitumen stabilised ballast particles hinder the geometric deterioration of railway track and ballast breakage [43, 44, 45], in detailed analysis it can be considered.

DEM simulations with particle flow code software (e.g. Itasca PFC^{3D}) can be useful in the future researches to be able to evaluate particle degradation. With this DEM method the expensive laboratory tests can be saved (if the model is validated with laboratory measurements), influence effect of lots of parameters can be considered, e.g. particle shape, PSD, stone physics, abraded particles, geosynthetic reinforcements, depth of ballast, etc. This method unfortunately very lengthy, and there is just a little chance for good results.

The authors' future aim is to utilize of the results and maybe to adopt these results of the research into national regulations, standards.

This article is the direct continuation of papers [46, 47] that are accepted manuscripts without publishing yet.

Acknowledgements

The publishing of this paper was supported by EFOP 3.6.1-16-2016-00017 project.

References

[1] Sz. Fischer, Crumbling examination of railway crushed stones by individual laboratory method, Sínek Világa, 57 (3) (2015) pp. 12–19, in Hungarian.

[2] Sz. Fischer, Breakage test of railway ballast materials with new laboratory method, Periodica Polytechnica Civil Engineering 61 (4) (2017) pp. 794–802. https://doi.org/10.3311/PPci.8549

[3] Sz. Fischer, A. Németh, D. Harrach, E. Juhász, Laboratory fatigue degradation tests of railway ballast materials, in: G. Köllő (Ed.), XXII. Conference on Civil Engineering and Architecture, Sumuleu Ciuc, 2018, pp. 58–61, in Hungarian.

[4] Sz. Fischer, A. Németh, Individual rock physics investigations of railway ballast materials, in: XI. Stone and Gravel Quarry Days, Velence, 2018, pp. 37–41, in Hungarian.

[5] Sz. Fischer, A. Németh, Special laboratory test for evaluation breakage (particle degradation) of railway ballast, in: Conference on Transport Sciences, Győr, 2018, pp. 87–96.

[6] E. Juhász, Sz. Fischer, Investigation of railway ballast materials' particle degradation with special laboratory test method, in: Abstract book of 14th Miklós Iványi International PhD & DLA Symposium, Pécs, 2018, pp. 89–90.

[7] Tests for mechanical and physical properties of aggregates. Part 1: Determination of the resistance to wear (micro-Deval), MSZ EN 1097-1:2012 (2012) in Hungarian.

[8] Tests for mechanical and physical properties of aggregates. Part 2: Methods for the determination of resistance to fragmentation, MSZ EN 1097-2:2010 (2010) in Hungarian.

[9] Aggregates for railway ballast, MSZ EN 13450:2003 (2003) in Hungarian.

[10] MÁV: Modification 4 in MÁV 102345/1995 Railway substructure and ballast quality acceptance regulations instruction (2010) pp. 1–5, in Hungarian.

[11] N. K. S. Al-Saoudi, K. H. Hassan, Behaviour of track ballast under repeated loading, Geotechnical and Geological Engineering 32 (1) (2014) pp. 167–178.

[12] S. C. Douglas, Ballast quality and breakdown during tamping, in: Joint Rail Conference, Knoxville, 2013, pp. 940–955.

[13] R. S. Kamalov, G. S. Ghataora, M. P. N. Burrow, M. Wehbi, P. Musgrave, in: Migration of fine particles from subgrade soil to the overlying ballast, in: Railway Engineering Conference, Edinburgh, 2017, pp 1–9.

[14] B. Indraratna, S. Nimbalkar, D. Christie, The performance of rail track incorporating the effects of ballast breakage, confining pressure and geosynthetic reinforcement, in: E. Tutumluer E., I. Al-Qadi (Eds.) Bearing Capacity of Roads, Railways and Airfields, Taylor and Frances, London, 2009, pp. 5–24.

[15] B. Indraratna, Y. Sun, S. Nimbalkar, Laboratory assessment of the role of particle size distribution on the deformation and degradation of ballast under

cyclic loading, Journal of Geotechnical and Geoenvironmental Engineering 142 (7) (2016) pp. 1–14.
https://doi.org/10.1061/(ASCE)GT.1943-5606.0001463

[16] A. Kolos, A. Konon, P. Chistyakov, Change of ballast strength properties during particle abrasive wear, Procedia Engineering 189 (2017) pp. 908–915.
https://doi.org/10.1016/j.proeng.2017.05.141

[17] V. Kondratov, V. Solovyova, I. Stepanova, The development of a high performance material for a ballast layer of a railway track, Procedia Engineering 189 (2017) pp. 823–828.
https://doi.org/10.1016/j.proeng.2017.05.128

[18] G. Liu, G. Jing, D. Ding, X. Shi, Micro-analysis of ballast angularity breakage and evolution by monotonic triaxial tests, in: X. Bian, Y. Chen, X. Ye (Eds.), Environmental Vibrations and Transportation Geodynamics, Springer, Singapore, 2018, pp. 133–144.
https://doi.org/10.1007/978-981-10-4508-0_12

[19] Y. Sun, C. Chen, S. Nimbalkar, Identification of ballast grading for rail track, Journal of Rock Mechanics and Geotechnical Engineering 9 (5) (2017) pp. 945–954.
https://doi.org/10.1016/j.jrmge.2017.04.006

[20] G. McDowell, Performance of geogrid-reinforced ballast, Ground Engineering January (2006), pp. 2–6.
URL https://www.geplus.co.uk/Journals/2014/06/20/b/q/m/GE-Jan-2006-Performance-of-geogrid-reinforced-ballast-McDowell-Stickley.pdf

[21] S. Nimbalkar, B. Indraratna, Field assessment of ballasted rail-roads using geosynthetics and shock mats, Procedia Engineering 143 (2016) pp. 1485–1494.
https://doi.org/10.1016/j.proeng.2016.06.175

[22] J. H. Xiao, D. Zhang, Y. H. Wang, Z. Luo, Cumulative deformation characteristic and shakedown limit of railway ballast under cyclic loading, in: The 10th International Conference on the Bearing Capacity of Roads, Railways and Airfields, Athens, 2017, pp. 1899–1904.

[23] Bearing capacity test on pavement structures, Plate bearing test, MSZ 2509-3:1989 (1989) in Hungarian.

[24] A. Danesh, M. Palassi, A. A. Mirghasemi, Evaluating the influence of ballast degradation on its shear behaviour, International Journal of Rail Transportation 6 (3) (2018) pp. 145–162. https://doi.org/10.1080/23248378.2017.1411212

[25] M. Gálos, L. Kárpáti, D. Szekeres, Railway ballast aggregates (Part 2), Sínek Világa 54 (1) (2011) pp. 6–13, in Hungarian.

[26] B. Lichtberger, Track compendium, Eurailpress, Hamburg, 2011.

[27] D. M. Kurhan, To the solution of problems about the railways calculation for strength taking into account unequal elasticity of the subrail base, Nauka ta Progres Transportu 55 (1) (2015) pp. 90–99.

[28] D. Kurhan, Determination of load for quasi-static calculations of railway track stress-strain state, Acta Technica Jaurinensis 9 (1) (2016) pp. 83–96. https://doi.org/10.14513/actatechjaur.v9.n1.400

[29] V. V. Kovalchuk, M. P. Sysyn, J. Sobolevska, O. Nabochenko, B. Parneta, A. Pentsak, Theoretical study into efficiency of the improved longitudinal profile of frogs at railroad switches, Eastern European Journal of Enterprise Technologies 4 (1) (2018) pp. 27–36. https://doi.org/10.15587/1729-4061.2018.139502

[30] M. P. Sysyn, V. V. Kovalchuk, D. Jiang, Performance study of the inertial monitoring method for railway turnouts, International Journal of Rail Transportation 4 (2018) pp. 33–42. https://doi.org/10.1080/23248378.2018.1514282

[31] M. P. Sysyn, U. Gerber, V. Kovalchuk, O. Nabochenko, The complex phenomenological model for prediction of inhomogeneous deformations of railway ballast layer after tamping works, Archives of Transport 46 (3) (2018) pp. 91–107. https://doi.org/10.5604/01.3001.0012.6512

[32] Á. Vinkó, Monitoring and condition assessment of tramway track using in-service vehicle, Pollack Periodica 11 (3) (2016) pp. 73–82. https://doi.org/10.1556/606.2016.11.3.7

[33] R. Nagy, Description of rail track geometry deterioration process in Hungarian rail lines No. 1 and No. 140, Pollack Periodica 12 (3) (2017) pp. 141–156.
https://doi.org/10.1556/606.2017.12.3.13

[34] I. D. Qunitanilla. Multi-scale study of the degradation of railway ballast, thesis, Mechanical engineering, Communauté Université Grenoble Alpes (2018).
URL https://tel.archives-ouvertes.fr/tel-01858650

[35] Y. L. Guo, G. Q. Jing, Ballast degradation analysis by Los Angeles Abrasion test and image analysis method, in: Loizos et al. (Eds.), The 10th International Conference on the Bearing Capacity of Roads, Railways and Airfields (BCRRA 2017), Athens, 2017, pp. 1811–1815.

[36] Y. Qian, H. Boler, M. Moaveni, E. Tutumluer, Y. M. A. Hashash, J. Ghaboussi, Characterizing Ballast Degradation through Los Angeles Abrasion Test and Image Analysis, Trasportation Research Record 2448 (1) (2014) pp. 142–151.
https://doi.org/10.3141/2448-17

[37] Y. Guo, V. Markine, J. Song, G. Jing, Ballast degradation: Effect of particle size and shape using Los Angeles Abrasion test and image analysis, Construction and Building Materials 169 (2018) pp. 414–424.
https://doi.org/10.1016/j.conbuildmat.2018.02.170

[38] Y. Qian, E. Tutumluer, D. Mishra, H. Kazmee, Behavior of Geogrid Reinforced Ballast at Different Levels of Degradation, Ground Improvement and Geosynthetics, in: Selected Papers from the Proceedings of the 2014 GeoShanghai International Congress, 2014 GeoShanghai International Congress: Ground Improvement and Geosynthetics - Shanghai, 2014, pp. 333–342.
https://doi.org/10.1061/9780784413401.033

[39] C. Ngamkhanong, S. Kaewunruen, C. Baniotopoulos, A review on modelling and monitoring of railway ballast, Structural Monitoring and Maintenance 4 (3) (2017) pp. 195–220.
https://doi.org/10.12989/smm.2017.4.3.195

[40] Y. Guo, V. Markine, J. Song, G. Jing, Ballast degradation: Effect of particle size and shape using Los Angeles Abrasion test and image analysis,

Construction and Building Materials 169 (2018) pp. 414–424. https://doi.org/10.1016/j.conbuildmat.2018.02.170

[41] Y. Guo, V. Markine, X. Zhang, W. Qiang, G. Jing, Image analysis for morphology, rheology and degradation study of railway ballast: A review, Transportation Geotechnics 18 (2019) pp. 173–211. https://doi.org/10.1016/j.trgeo.2018.12.001

[42] E. Salvatore, G. Modoni, E. Ando, M. Albano, G. Viggiani, Determination of the critical state of granular materials with triaxial tests, Soils and Foundations 57 (5) (2017) pp. 733–744. https://doi.org/10.1016/j.sandf.2017.08.005

[43] M. Sol-Sánchez, G. D'Angelo, Review of the design and maintenance technologies used to decelerate the deterioration of ballasted railway tracks, Construction and Building Materials 157 (2017) pp. 402–415. https://doi.org/10.1016/j.conbuildmat.2017.09.007

[44] M. Giunta, S. Bressi, G. D'Angelo, Life Cycle Cost Assessment of Bitumen Stabilised Ballast: a novel maintenance strategy for railway track-bed, Construction and Building Materials 172 (2018) pp. 751–759. https://doi.org/10.1016/j.conbuildmat.2018.04.020

[45] G. D'Angelo, S. Bressi, M. Giunta, D. Lo Presti, N. Thom, Novel Performance-Based Technique for Predicting Maintenance Strategy of Bitumen Stabilised Ballast, Construction and Building Materials 161 (2018) pp. 1–8. https://doi.org/10.1016/j.conbuildmat.2017.11.115

[46] E. Juhász, Sz. Fischer, Breakage tests of railway ballast stone materials with using of laboratory dynamic pulsating, Sínek Világa 61 (1) (2019) pp. 16–21, in Hungarian.

[47] E. Juhász, Sz. Fischer, Investigation of railroad ballast particle breakage, accepted manuscript without publishing yet, Pollack Periodica 14 (2019).

Acta
Technica
Jaurinensis

Vol.12, No.1, pp. 55–81, 2019
DOI: 10.14513/actatechjaur.v12.n1.492
Available online at acta.sze.hu

Comprehensive Survey of PID Controller Design for the Inverted Pendulum

M. Kuczmann

Széchenyi István University, Department of Automation
Egyetem tér 1, H-9026, Győr, Hungary
E-mail: kuczmann@sze.hu

Abstract: The survey shows the detailed PID controller design how to stabilize the inclination angle as well as the horizontal movement of an inverted pendulum on a cart system step by step. Pendulum model is based on Euler-Lagrange modeling, and the nonlinear state space model is linearized in the unstable upward position. Controller design is performed by applying the transfer function description. The pendulum has been inserted into a virtual reality laboratory, which is suitable to use in model based control teaching.

Keywords: *inverted pendulum, model based control, PID controller design*

1. Introduction

The paper presents a comprehensive study of PID controller design for an inverted pendulum mounted on a cart.

The inverted pendulum is an unstable system that must be stabilized by the pushing-pulling force $F = F(t)$ acting on the cart (Fig. 1), i.e. to reach $\varphi(t) \to 0$ in the stationary state, where $\varphi = \varphi(t)$ is the inclination angle. The pendulum simply falls over if the cart is not moved to balance it. The actuator is typically an electric motor. The cart can move only along the x-axis.

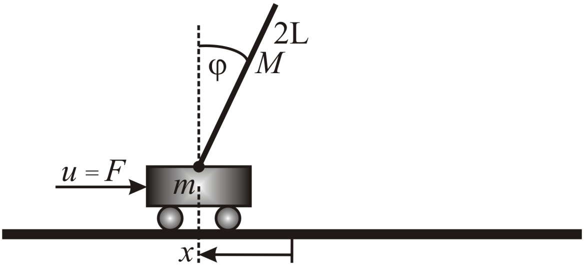


Figure 1. The inverted pendulum model.

The system input is the force, $u(t) = F(t)$. First, the output is only the inclination angle φ resulting a single input single output system ($y(t) = \varphi(t)$). Second, the horizontal movement $x = x(t)$ is also stabilized, i.e. the plant is a single input and multiple output system ($y_1(t) = x(t)$, $y_2(t) = \varphi(t)$).

The studied plant is a popular example commonly found in control system textbooks and research literature [1–7]. The dynamics of the system are nonlinear as presented in the paper based on the above mentioned literature, but PID controller design is based on the linearized system. All the PID type controllers are presented in detail.

The aim of this survey paper is to show the Euler-Lagrange modeling of the inverted pendulum system, than the linear PID controller design step by step. The mentioned formulations are deeply presented in detail, that it why it can be used in teaching of model based control.

The really operating device has not built in this research, however the virtual reality based implementation has been performed which is applicable to understand the steps of control design.

A real-world example that relates directly to this inverted pendulum system is the attitude control of a booster rocket at takeoff or the well-known personal transporter Segway.

2. Dynamic model of the pendulum

To set up the dynamic model of pendulum, the Euler-Lagrange equation is applied,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial K}{\partial \dot{q}_i} - \frac{\partial K}{\partial q_i} + \frac{\partial P}{\partial \dot{q}_i} = \tau_i, \tag{1}$$

where K is the kinetic energy, P is the potential energy, q_i and τ_i are the generalized coordinate and the generalized torque (force), respectively. In the case of pendulum $i = 1{,}2$, i.e. $q_1 = x$ and $q_2 = \varphi$, moreover $\tau_1 = F$ and $\tau_2 = 0$. Here F is the force pulling or pushing the cart. This is an underactuated system because there are two output signals (x and φ) and only one input signal (F).

The kinetic energy of the system is as follows:

$$K = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}Mv_M^2 + \frac{1}{2}\Theta\dot{\varphi}^2, \tag{2}$$

with the mass of cart, m, the mass of rod, M, and the inertial moment of the rod belonging to the center of mass, Θ. The value of the last term is $\Theta = \frac{1}{3}ML^2$ (the length of the rod is $2L$). The coordinates and the velocity of the center of mass of the rod are x_M, y_M and v_M.

The potential energy of the system is

$$P = MgLC_\varphi, \tag{3}$$

where g is the gravitational acceleration. For simplicity, $S_\varphi = \sin\varphi$ and $C_\varphi = \cos\varphi$ notations are used in the paper.

The velocity of the rod center of mass can be obtained by the coordinates as follows:

$$\begin{aligned} v_M^2 = \dot{x}_M^2 + \dot{y}_M^2 &= \left(\dot{x} + (LS_\varphi)'\right)^2 + \left((LC_\varphi)'\right)^2 \\ &= \dot{x}^2 + 2LC_\varphi\dot{x}\dot{\varphi} + L^2\dot{\varphi}^2, \end{aligned} \tag{4}$$

since $x_M = x + LS_\varphi$ and $y_M = LC_\varphi$.

The terms in (1) can be obtained easily:

$$\begin{aligned} &\frac{\partial K}{\partial \dot{x}} = m\dot{x} + M\dot{x} + MLC_\varphi\dot{\varphi}, && \frac{\partial K}{\partial x} = 0, && \frac{\partial P}{\partial x} = 0, \\ &\frac{\partial K}{\partial \dot{\varphi}} = MLC_\varphi\dot{x} + ML^2\dot{\varphi} + \Theta\dot{\varphi}, && \frac{\partial K}{\partial \varphi} = -MLS_\varphi\dot{x}\dot{\varphi}, && \frac{\partial P}{\partial \varphi} = -MgLS_\varphi, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial K}{\partial \dot{x}} &= m\ddot{x} + M\ddot{x} + MLC_\varphi\ddot{\varphi} - MLS_\varphi\dot{\varphi}^2, \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial K}{\partial \dot{\varphi}} &= MLC_\varphi\ddot{x} - MLS_\varphi\dot{x}\dot{\varphi} + ML^2\ddot{\varphi} + \Theta\ddot{\varphi}. \end{aligned} \tag{6}$$

Putting everything together gives the following differential equations:

$$\begin{aligned} &(m+M)\ddot{x} + MLC_\varphi\ddot{\varphi} - MLS_\varphi\dot{\varphi}^2 = F, \\ &MLC_\varphi\ddot{x} + (ML^2+\Theta)\ddot{\varphi} - MgLS_\varphi = 0. \end{aligned} \tag{7}$$

From these equations, after some algebraic manipulations, the second order derivatives can be yielded as

$$\ddot{\varphi} = \frac{MLS_\varphi C_\varphi\dot{\varphi}^2 - (m+M)gS_\varphi + FC_\varphi}{MLC_\varphi^2 - \frac{4}{3}(m+M)L}, \tag{8}$$

and

$$\ddot{x} = \frac{\frac{4}{3}MLS_\varphi\dot{\varphi}^2 - MgS_\varphi C_\varphi + \frac{4}{3}F}{\frac{4}{3}(m+M) - MC_\varphi^2}. \tag{9}$$

The two second order differential equations can be rewritten as four first order differential equations by introducing state variables: $x_1 = x$, $x_2 = \dot{x}$, $x_3 = \varphi$, $x_4 = \dot{\varphi}$, i.e. $x_2 = \dot{x}_1$ and $x_4 = \dot{x}_3$. Finally, the state space representation of the dynamic modell is the following:

$$\begin{aligned}
\dot{x}_1 &= x_2,\\
\dot{x}_2 &= \frac{\frac{4}{3}MLS_{x_3}\dot{x}_4^2 - MgS_{x_3}C_{x_3} + \frac{4}{3}F}{\frac{4}{3}(m+M) - MC_{x_3}^2},\\
\dot{x}_3 &= x_4,\\
\dot{x}_4 &= \frac{MLS_{x_3}C_{x_3}\dot{x}_4^2 - (m+M)gS_{x_3} + FC_{x_3}}{MLC_{x_3}^2 - \frac{4}{3}(m+M)L}.
\end{aligned} \tag{10}$$

This nonlinear system can be linearized in the unstable upright position, when $\varphi = 0$ and $\dot{\varphi} = 0$, i.e. the approximations $S_\varphi \cong \varphi$ and $C_\varphi \cong 1$ can be applied.

At the end, the linearized system can be modelled by the following state space equations:

$$\begin{aligned}
\dot{x}_1 &= x_2,\\
\dot{x}_2 &= \frac{-3Mg}{4m+M}x_3 + \frac{4}{4m+M}F,\\
\dot{x}_3 &= x_4,\\
\dot{x}_4 &= \frac{3(m+M)g}{4mL+ML}x_3 - \frac{3}{4mL+ML}F.
\end{aligned} \tag{11}$$

It can be written in the usual matrix form of SISO (single input single output) systems as

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{b}u,\\
y &= \mathbf{c}^{\mathrm{T}}\mathbf{x} + Du,
\end{aligned} \tag{12}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0\\ 0 & 0 & \frac{-3Mg}{4m+M} & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & \frac{3(m+M)g}{4mL+ML} & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0\\ \frac{4}{4m+M}\\ 0\\ -\frac{3}{4mL+ML} \end{bmatrix}. \tag{13}$$

Let the output y of the system be the variable φ, in this case

$$\mathbf{c}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}, \quad D = 0. \tag{14}$$

Here $\mathbf{x} = [x_1\, x_2\, x_3\, x_4]^{\mathrm{T}}$ is the vector of the state variables.

In the case of the two output system $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}u$, where

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{15}$$

and $\mathbf{y} = [x\, \varphi]^{\mathrm{T}}$ is the output vector.

For simplicity, the following notations are used to obtain the transfer function of the model:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & q & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & p^2 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ \beta \\ 0 \\ \alpha \end{bmatrix}. \tag{16}$$

The transfer function of the system (13)-(14) with the notations defined by (16) is the following:

$$W_{\mathrm{P}}(s) = \frac{\Phi(s)}{U(s)} = \frac{\mathcal{L}\{\varphi(t)\}}{\mathcal{L}\{u(t)\}} = \frac{\alpha}{s^2 - p^2}. \tag{17}$$

Here, the subscript P is for the plant, and the operator $\mathcal{L}\{\}$ represents the Laplace transform.

The transfer function

$$W_{\mathrm{P}}(s) = \frac{-0.01}{s^2 - 1.47} = \frac{-0.01}{(s + 1.21)(s - 1.21)} \tag{18}$$

is applied in the paper to represent the results ($m = 5$, $M = 10$, $2L = 20$ in a coherent unit system, furthermore $\alpha = -0.01$, $p = 1.21$, $\beta = 0.13$, $q = -9.81$).

It is easy to see that the system is unstable, because one of the poles is surely positive: $p_{1,2} = \pm p$. It is noted that the four eigenvalues of the system matrix $\mathbf{A}$ are $\lambda_{1,2} = 0$, $\lambda_{3,4} = \pm p$.

3. Stabilizing the SISO plant by PID controllers

The typical closed loop block diagram of the controller system is shown in Fig. 2, where the Laplace transform of the signals are highlighted: the input and output of the plant are $U = U(s)$ and $\Phi = \Phi(s)$, the reference signal is zero, and $E = E(s) = 0 - \Phi$ is the error.

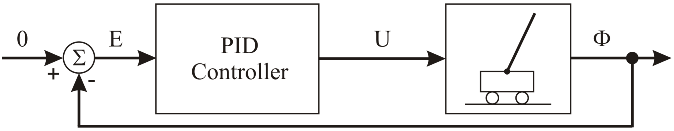


Figure 2. Block diagram of the controller system.

The transfer function of the open loop is $W_\mathrm{O} = W_\mathrm{C} W_\mathrm{P}$, where W_C and W_P are the transfer function of the controller and the plant, respectively. The notation (s) will be cancelled in the following part of the paper.

In the followings, all the four controllers are studied.

3.1. Applying P controller

In the case of P controller, W_C is a constant, denoted by K_P ($W_\mathrm{C} = K_\mathrm{P}$), furthermore

$$W_\mathrm{O} = K_\mathrm{P} \frac{\alpha}{s^2 - p^2}. \qquad (19)$$

First, the Nyquist criterion and Nyquist plot are used to check whether the plant stabilization can be performed or not. It is well known that the Nyquist contour of this open loop system (with one unstable pole) should encircle counter clock-wise the point $-1 + \mathrm{j}0$ once. It can not be satisfied as it is demonstrated in Fig. 3 when the gain is negative (e.g. $K_\mathrm{P} = -200$). The Nyquist contour encircles the point $-1 + \mathrm{j}0$ once, but its direction is clock-wise. When K_P is positive, the Nyquist plot is on the right hand side plane resulting non stable system.

The result can be verified by the root locus of the open loop transfer function (Fig. 3). One of the poles stands on the right hand side of the complex plane while $K_{\mathrm{P}} \geq 1$ is changing. It is not shown here, but the poles become unstable conjugate complex pairs when K_{P} is negative and large enough.

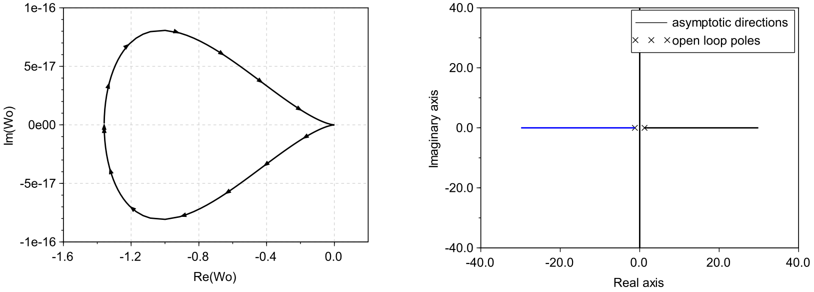


Figure 3. Nyquist plot and root locus of open loop system with P controller.

Now, let us prove this by analytically checking the poles of the closed loop system, too. The transfer function of the closed loop system is

$$W_{\mathrm{CL}} = \frac{W_{\mathrm{O}}}{1 + W_{\mathrm{O}}} = \frac{B}{A + B}, \tag{20}$$

if $W_{\mathrm{O}} = \frac{B}{A}$, moreover $B = B(s)$ and $A = A(s)$ are the numerator and the denominator of the open loop transfer function, respectively. The roots of the polynomial $A + B$ (the poles of W_{CL}) are responsible for the stability of the closed loop system.

In the case of P controller $A + B = s^2 - p^2 + K_{\mathrm{P}}\alpha$, from which the poles are $p_{1,2} = \pm\sqrt{p^2 - K_{\mathrm{P}}\alpha}$. If K_{P} is positive, one of the poles is usually unstable ($\alpha < 0$). If K_{P} is negative, the poles become complex numbers with zero real part.

At the end, it is concluded that stabilization can not be performed by simple P controller.

3.2. Applying PD controller

In the case of PD controller,

$$W_{\mathrm{C}} = K_{\mathrm{PD}} \frac{1 + sT_{\mathrm{D}}}{1 + sT'_{\mathrm{D}}}, \tag{21}$$

where T_{D} and T'_{D} are time constants, and K_{PD} is the gain of the controller.

First of all, W_{P} is rewritten in the form

$$W_{\mathrm{P}} = \frac{\alpha}{s^2 - p^2} = \frac{-\frac{\alpha}{p^2}}{\left(1 + \frac{s}{p}\right)\left(1 - \frac{s}{p}\right)}. \tag{22}$$

Next, the stable pole of the plant is cancelled by T_{D}, i.e. $T_{\mathrm{D}} = \frac{1}{p}$. The open loop transfer function becomes

$$W_{\mathrm{O}} = -\frac{K_{\mathrm{PD}}\alpha}{p^2} \frac{1}{(1 + sT'_{\mathrm{D}})\left(1 - \frac{s}{p}\right)}. \tag{23}$$

The following inequality must be satisfied when selecting the value of T'_{D}:

$$\frac{1}{T'_{\mathrm{D}}} > \frac{1}{T_{\mathrm{D}}}. \tag{24}$$

The root locus of the open loop is shown first in Fig. 4 with positive and negative gain (here $T'_{\mathrm{D}} = T_{\mathrm{D}}/10$ is used). It is easy to see that the system can not be stabilized by any positive gain, but appropriate negative gain can stabilize the pendulum.

It can be checked by the Nyquist diagram of the open loop transfer function as well. See Fig. 5 for $K_{\mathrm{PD}} = -200$. The Nyquist contour encircles the point $-1 + \mathrm{j}0$ once counter clock-wise if the value of K_{PD} is large enough with negative sign.

Fig. 6 shows the impulse response of the closed loop system. The transient can oscillate when the gain is too high, anyway the stabilization time is shorter.

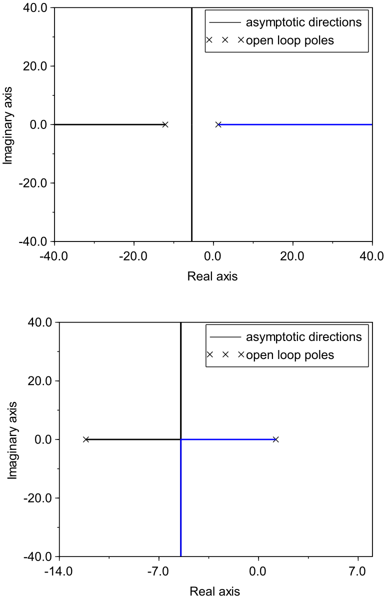


Figure 4. Open loop root locus of PD controller with positive and negative gain.

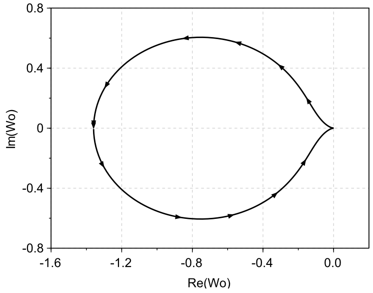


Figure 5. Nyquist plot of open loop system with PD controller.

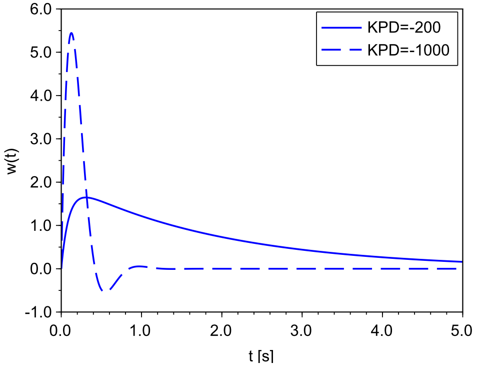


Figure 6. Impulse response of the closed loop system.

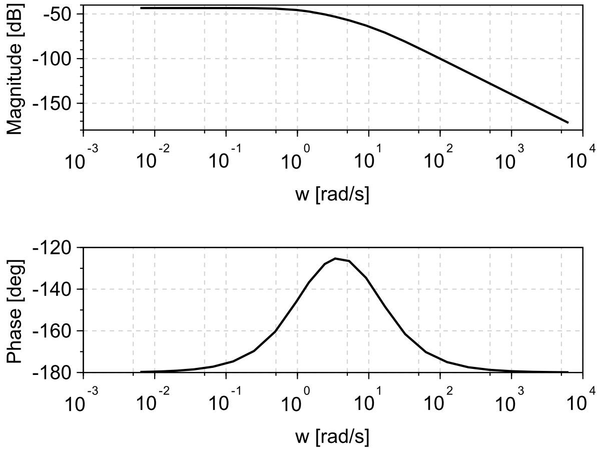


Figure 7. Bode plot of open loop system to determine gain.

The open loop Bode plot is shown in Fig. 7, from which the maximum phase margin can be read easily, $PM_{\max} \cong 55°$ at $\omega_{\mathrm{c}} \cong 3.8$, i.e. $K_{\mathrm{PD}} = -456$. In this case, there is no oscillation.

It can be computed analytically, too. The phase of the open loop system is the following:

$$\Phi(\omega) = -180° - \operatorname{atan} \omega T'_{\mathrm{D}} + \operatorname{atan} \frac{\omega}{p} \tag{25}$$

having the maximum value of phase margin at $\omega = \sqrt{\frac{p}{T'_{\mathrm{D}}}}$, i.e. the maximum phase margin is

$$PM_{\max} = -180° - \operatorname{atan} \sqrt{pT'_{\mathrm{D}}} + \operatorname{atan} \frac{1}{\sqrt{pT'_{\mathrm{D}}}}. \tag{26}$$

By these nonlinear expressions, the phase margin and the cut-off frequency, i.e. the transient behavior, can be designed semi-analytically.

For example, the PD controller

$$W_{\mathrm{C}} = -456 \frac{1 + s0.82}{1 + s0.082} \tag{27}$$

is a good choice to stabilize this system with maximum phase margin without oscillation.

Finally, the Routh-Hurwitz criterion is applied. In the case of PD controller $A + B = s^2 T'_{\mathrm{D}} p + s(p - p^2 T'_{\mathrm{D}}) + (K_{\mathrm{PD}}\alpha - p^2)$ which has the same roots as the monic polynomial $s^2 + s\frac{(1-pT'_{\mathrm{D}})}{T'_{\mathrm{D}}} + \frac{(K_{\mathrm{PD}}\alpha - p^2)}{T'_{\mathrm{D}} p}$. According to the Routh-Hurwitz criterion, the coefficients of the above polynomials, after inserting the designed controller parameters must be positive.

PD controller can be designed by the Routh-Hurwitz criterion as well by prescribing the coefficients of the characteristic polynomial, i.e. by setting the value of p_1 and p_0,

$$s^2 + s \underbrace{\frac{(1 - pT'_{\mathrm{D}})}{T'_{\mathrm{D}}}}_{p_1} + \underbrace{\frac{(K_{\mathrm{PD}}\alpha - p^2)}{T'_{\mathrm{D}} p}}_{p_0} = 0, \tag{28}$$

i.e.

$$T_\mathrm{D} = \frac{1}{p}, \quad T'_\mathrm{D} = \frac{1}{p + p_1}, \quad K_\mathrm{PD} = \frac{p^2 + p_0 p T'_\mathrm{D}}{\alpha}. \tag{29}$$

An example is presented here to show how to use these expressions. Let us, say, determine the overshoot of any transient to be less then $\Delta v_\mathrm{max} = 5\%$ and the settling time to be $T_\mathrm{s} = 2$ with the tolerance fraction $\Delta = 2\%$. These criteria can be represented by supposing the dominant pole pair of the closed loop system, $p_{1,2} = -\xi\Omega \pm \mathrm{j}\Omega\sqrt{1-\xi^2}$, where ξ and Ω are the damping ratio and the natural frequency of the complex pole pair. The damping ratio and the real part should be higher than ($\sigma = \xi\Omega$)

$$\xi_\mathrm{min} = \frac{1}{\sqrt{1 + \frac{\pi^2}{\ln^2 \Delta v_\mathrm{max}}}} \cong 0.7, \quad \text{and} \quad \sigma_\mathrm{min} = -\frac{\ln\Delta}{T_\mathrm{s}} \cong 2. \tag{30}$$

The following dominant pole pair, for example, can satisfy these criteria ($\xi = 0.9$, $\sigma = \xi\Omega = 5$): $p_{1,2} = -5 \pm \mathrm{j}2.4$, i.e. $p_1 = 10$ and $p_0 = 30.8$, and

$$W_\mathrm{C} = -480\frac{1 + s0.82}{1 + s0.089}. \tag{31}$$

It can be concluded that the pendulum can be stabilized by a PD controller.

3.3. Applying PI controller

The transfer function of the PI controller is the following:

$$W_\mathrm{C} = K_\mathrm{PI}\frac{1 + sT_\mathrm{I}}{sT_\mathrm{I}}, \tag{32}$$

where T_I and K_PI are the time constant and the gain of the controller, respectively.

The stable pole of the plant is cancelled by T_I, i.e. $T_\mathrm{I} = \frac{1}{p}$. The open loop transfer function can be written as

$$W_\mathrm{O} = -\frac{K_\mathrm{PI}\alpha}{p^2}\frac{1}{sT_\mathrm{I}\left(1 - \frac{s}{p}\right)}. \tag{33}$$

It is easy to check that $A + B = s^2 - sp + K_{\mathrm{PI}}\alpha$. The sign of the second term $-sp$ is usually negative, because $p > 0$. It means that the Routh-Hurwitz criterion can not be satisfied, and the plant can not be stabilized by PI controller.

3.4. Applying PID controller

The following transfer function of PID controller has been used in this paper:

$$W_{\mathrm{C}} = K_{\mathrm{PID}} \frac{1 + sT_{\mathrm{I}}}{sT_{\mathrm{I}}} \frac{1 + sT_{\mathrm{D}}}{1 + sT'_{\mathrm{D}}}. \tag{34}$$

The stable pole of the plant is cancelled by T_{D}, i.e. $T_{\mathrm{D}} = \frac{1}{p}$. The time constant of the integrator can also be set to the value $T_{\mathrm{I}} = \frac{1}{p}$. The inequality (24) must be satisfied again when selecting the value of T'_{D}. Finally, the open loop transfer function can be written as

$$W_{\mathrm{O}} = -\frac{K_{\mathrm{PID}}\alpha}{p^2} \frac{1 + \frac{s}{p}}{\frac{s}{p}} \frac{1}{(1 + sT'_{\mathrm{D}})\left(1 - \frac{s}{p}\right)}. \tag{35}$$

Fig. 8 shows the root locus of the open loop system with negative gain when

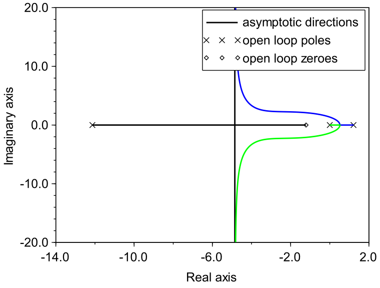


Figure 8. Root locus of open loop system with PID controller with negative gain.

$T_\mathrm{D}' = T_\mathrm{D}/10$. It can be seen that the pendulum can be stabilized by a high enough negative gain. With positive gain it is not possible.

The maximum phase margin has been determined by the Bode plot resulting the gain $K_\mathrm{PID} = -871$ and the cut-off frequency $\omega_\mathrm{c} \cong 5.8$.

The open loop Nyquist diagram can be seen in Fig. 9. This is a magnified plot to check the counter clock-wise direction of the contour around the point $-1 + \mathrm{j}0$.

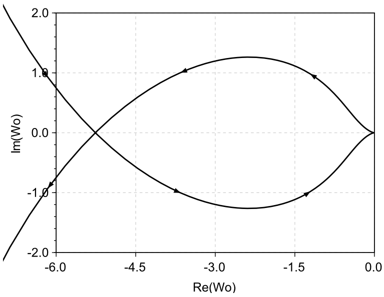


Figure 9. The open loop Nyquist diagram with PID controller, $K_\mathrm{PID} = -871$.

The PID controller

$$W_\mathrm{C} = -871 \frac{1 + s0.82}{s0.82} \frac{1 + s0.82}{1 + s0.082} \tag{36}$$

is a good candidate to stabilize the system resulting the closed loop impulse response shown in Fig. 10.

Finally, the Routh-Hurwitz criterion is applied to check the stability. The following polynomial is obtained:

$$A + B = s^3 \underbrace{T_\mathrm{D}' p}_{a_3} + s^2 \underbrace{\left(p - p^2 T_\mathrm{D}'\right)}_{a_2} + s \underbrace{\left(K_\mathrm{PID}\alpha - p^2\right)}_{a_1} + \underbrace{K_\mathrm{PID}\alpha p}_{a_0} \tag{37}$$

Not only the coefficients a_0, a_1, a_2 and a_3 must be positive but there is one more criterion according to the Routh-Hurwitz rule: $a_1 a_2 - a_0 a_3$ must be positive.

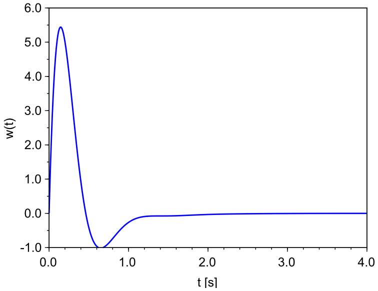


Figure 10. Impulse response of the closed loop system.

The controller can be designed by prescribing the coefficients of the characteristic polynomial, i.e. by setting the value of p_2, p_1 and p_0, moreover T_I is free,

$$s^3 + s^2 \underbrace{\frac{1 - pT'_\text{D}}{T'_\text{D}}}_{p_2} + s \underbrace{\frac{K_\text{PID}\alpha - p^2}{pT'_\text{D}}}_{p_1} + \underbrace{\frac{K_\text{PID}\alpha}{pT_\text{I}T'_\text{D}}}_{p_0} = 0, \tag{38}$$

from which the controller parameters can be obtained as

$$T_\text{D} = \frac{1}{p}, \quad T'_\text{D} = \frac{1}{p + p_2}, \quad K_\text{PID} = \frac{p^2 + p_1 T'_\text{D}}{\alpha}, \quad T_\text{I} = \frac{\alpha K_\text{PID}}{pp_0 T'_\text{D}}. \tag{39}$$

It can be concluded that PID controller is able to stabilize the pendulum.

4. Stabilizing the horizontal movement

Unfortunately, there is a problem with the above mentioned controllers: the cart moves along the x-axis, although the PD and the PID controllers are able to stabilize the angle of the pendulum. This design would not be feasible to implement on an actual physical system. In this Section not only the angle, but the horizontal movement is stabilized by PID controllers according to Fig. 11.

The controll signal $u(t)$ is the sum of two controller signals [5], i.e.

$$U = -W_{\mathrm{C}_1}\left(X_{\mathrm{r}} - X\right) + W_{\mathrm{C}_2}\left(0 - \Phi\right). \tag{40}$$

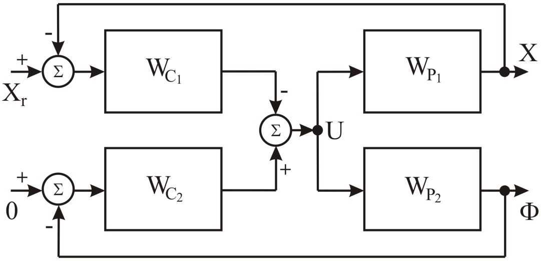


Figure 11. Block diagram of the two loop controller system.

The following transfer function with two integrators must be appended:

$$W_{\mathrm{P}_1} = \frac{X}{U} = \frac{\beta s^2 + \gamma}{s^2(s^2 - p^2)}, \tag{41}$$

where $\gamma = \alpha q - \beta p^2$. The horizontal movement is according to the integrators in this transfer function.

In this section the following notation will be used for (17):

$$W_{\mathrm{P}_2}(s) = \frac{\Phi}{U} = \frac{\alpha}{s^2 - p^2}. \tag{42}$$

The horizontal movement and the inclination angle can be expressed by the closed loop transfer functions,

$$X = \frac{-W_{\mathrm{P}_1}W_{\mathrm{C}_1}}{1 - W_{\mathrm{P}_1}W_{\mathrm{C}_1} + W_{\mathrm{P}_2}W_{\mathrm{C}_2}} X_{\mathrm{r}}, \tag{43}$$

and

$$\Phi = \frac{-W_{\mathrm{P}_2}W_{\mathrm{C}_1}}{1 - W_{\mathrm{P}_1}W_{\mathrm{C}_1} + W_{\mathrm{P}_2}W_{\mathrm{C}_2}} X_{\mathrm{r}}. \tag{44}$$

The closed loop system stability is depending on the denominator term $1 - W_{\mathrm{P}_1} W_{\mathrm{C}_1} + W_{\mathrm{P}_2} W_{\mathrm{C}_2}$. This means that the two controllers W_{C_1} and W_{C_2} can not be designed independently.

There are 16 possible combinations of the two controllers: P-P, P-PI, P-PD, P-PID, PI-P, PI-PI, PI-PD, PI-PID, PD-P, PD-PI, PD-PD, PD-PID, PID-P, PID-PI, PID-PD, PID-PID. Unfortunately, not all of these setups are feasible. Next, only the controller transfer functions mentioned in Section 3 are studied without analyzing the possible Diophantine equations [1, 2] (this is planned in a separate paper). It is noted that the time constants have no got physical meaning, they are just parameters of the controller.

4.1. P controller for horizontal movement

In the case of P-P configuration $W_{\mathrm{C}_1} = K_{\mathrm{P}}^1$, $W_{\mathrm{C}_2} = K_{\mathrm{P}}^2$, i.e.

$$1 - W_{\mathrm{P}_1} W_{\mathrm{C}_1} + W_{\mathrm{P}_2} W_{\mathrm{C}_2} = 1 - \frac{\beta s^2 + \gamma}{s^2(s^2 - p^2)} K_{\mathrm{P}}^1 + \frac{\alpha}{s^2 - p^2} K_{\mathrm{P}}^2 = 0. \tag{45}$$

After a short algebra, the following polynomial can be obtained:

$$s^4 + 0s^3 + \left(\alpha K_{\mathrm{P}}^2 - \beta K_{\mathrm{P}}^1 - p^2\right) s^2 + 0s + \left(-K_{\mathrm{P}}^1 \gamma\right), \tag{46}$$

resulting a non stable closed loops system, because the coefficients of the terms s^3 and s are zero.

It is easy to check the statement that, the P-PI system is similarly unstable. In this case $W_{\mathrm{C}_1} = K_{\mathrm{P}}^1$, $W_{\mathrm{C}_2} = K_{\mathrm{PI}}^2 \frac{1+sT_{\mathrm{I}}^2}{sT_{\mathrm{I}}^2}$, and the coefficient of the term s^3 is zero.

In the case of P-PD configuration $W_{\mathrm{C}_1} = K_{\mathrm{P}}^1$, $W_{\mathrm{C}_2} = K_{\mathrm{PD}}^2 \frac{1+sT_{\mathrm{D}}^2}{1+sT_{\mathrm{D}}'^2}$, i.e.

$$1 - W_{\mathrm{P}_1} W_{\mathrm{C}_1} + W_{\mathrm{P}_2} W_{\mathrm{C}_2} = 1 - \frac{\beta s^2 + \gamma}{s^2(s^2 - p^2)} K_{\mathrm{P}}^1 + \frac{\alpha}{s^2 - p^2} K_{\mathrm{PD}}^2 \frac{1 + sT_{\mathrm{D}}^2}{1 + sT_{\mathrm{D}}'^2} = 0, \tag{47}$$

from which, after a short algebra, the following polynomial can be given:

$$\begin{aligned} &s^5+\frac{1}{T_{\mathrm{D}}^{\prime 2}}s^4+\left(-p^2-\beta K_{\mathrm{P}}^1+\alpha K_{\mathrm{PD}}^2\frac{T_{\mathrm{D}}^2}{T_{\mathrm{D}}^{\prime 2}}\right)s^3+\frac{-p^2-\beta K_{\mathrm{P}}^1+\alpha K_{\mathrm{PD}}^2}{T_{\mathrm{D}}^{\prime 2}}s^2 \\ &\quad+\left(-\gamma K_{\mathrm{P}}^1\right)s+\left(-\frac{\gamma K_{\mathrm{P}}^1}{T_{\mathrm{D}}^{\prime 2}}\right). \end{aligned} \tag{48}$$

The P-PD configuration is redundant, because there are four parameters to be determined (K_{P}^1, K_{PD}^2, T_{D}^2, $T_{\mathrm{D}}^{\prime 2}$), but there are five equations according to the fifth order polynomial. In the followings, redundant systems will be skipped.

The P-PID system is finally feasible. From the characteristic equation

$$1-\frac{\beta s^2+\gamma}{s^2(s^2-p^2)}K_{\mathrm{P}}^1+\frac{\alpha}{s^2-p^2}K_{\mathrm{PID}}^2\frac{1+sT_{\mathrm{I}}^2}{sT_{\mathrm{I}}^2}\frac{1+sT_{\mathrm{D}}^2}{1+sT_{\mathrm{D}}^{\prime 2}}=0 \tag{49}$$

the following polynomial can be got:

$$\begin{aligned} &s^5+\underbrace{\frac{1}{T_{\mathrm{D}}^{\prime 2}}}_{p_4}s^4+\underbrace{\left(-p^2-\beta K_{\mathrm{P}}^1\right)}_{p_3}s^3+\underbrace{\frac{-p^2-\beta K_{\mathrm{P}}^1+\alpha K_{\mathrm{PID}}^2T_{\mathrm{D}}^2}{T_{\mathrm{D}}^{\prime 2}}s^2}_{p_2} \\ &\quad+\underbrace{\left(-\gamma K_{\mathrm{P}}^1+\alpha K_{\mathrm{PID}}^2\frac{T_{\mathrm{D}}^2}{T_{\mathrm{I}}^2T_{\mathrm{D}}^{\prime 2}}+\alpha K_{\mathrm{PID}}^2\frac{1}{T_{\mathrm{D}}^{\prime 2}}\right)}_{p_1}s+\underbrace{\frac{\alpha K_{\mathrm{PID}}^2-\gamma K_{\mathrm{P}}^1T_{\mathrm{I}}^2}{T_{\mathrm{I}}^2T_{\mathrm{D}}^{\prime 2}}}_{p_0}. \end{aligned} \tag{50}$$

From the resulting five equations the following controller parameters can be obtained analytically:

$$T_{\mathrm{D}}^{\prime 2}=\frac{1}{p_4},\quad K_{\mathrm{P}}^1=-\frac{p_0T_{\mathrm{D}}^{\prime 2}}{\gamma}, \tag{51}$$

and T_{D}^2 is the solution of the second order equation

$$\left(p_1+\gamma K_{\mathrm{P}}^1\right)T_{\mathrm{D}}^{\prime 2}\left(T_{\mathrm{D}}^2\right)^2+\left(-p^2-\beta K_{\mathrm{P}}^1-p_2T_{\mathrm{D}}^{\prime 2}\right)T_{\mathrm{D}}^2+\left(p_3+p^2+\beta K_{\mathrm{P}}^1\right)T_{\mathrm{D}}^{\prime 2}=0, \tag{52}$$

finally

$$T_{\mathrm{I}}^2=\frac{p_3+p^2+\beta K_{\mathrm{P}}^1}{\left(p_1+\gamma K_{\mathrm{P}}^1\right)T_{\mathrm{D}}^2},\quad K_{\mathrm{PID}}^2=\frac{p_1+\gamma K_{\mathrm{P}}^1}{\alpha}T_{\mathrm{I}}^2T_{\mathrm{D}}^{\prime 2}. \tag{53}$$

The impulse response of this system can be seen in Fig. 12 when all the desired poles of the closed loop system are equal to -1, i.e. the characteristic polynomial is prescribed by $s^5 + 5s^4 + 10s^3 + 10s^2 + 5s + 1$. It can be seen that all the system outputs have been stabilized. The behavior of the closed loop system can be set by the desired poles.

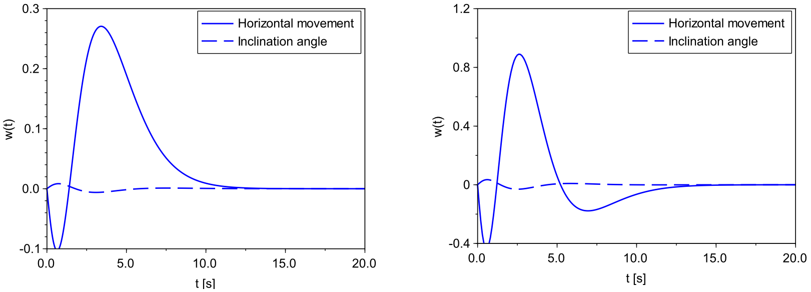


Figure 12. Impulse response of the closed loop P-PID and PI-PID systems.

4.2. PI controller for horizontal movement

Only PI-PID can be used as stabilizing controller from the second group. It is easy to check that, configurations PI-P, PI-PI and PI-PD result in redundant systems, moreover some of the polynomial coefficients are equal to zero.

The design of a feasible PI-PID system can be performed by the characteristic equation

$$1 - \frac{\beta s^2 + \gamma}{s^2(s^2 - p^2)} K_{\mathrm{PI}}^{1} \frac{1 + sT_{\mathrm{I}}^{1}}{sT_{\mathrm{I}}^{1}} + \frac{\alpha}{s^2 - p^2} K_{\mathrm{PID}}^{2} \frac{1 + sT_{\mathrm{I}}^{2}}{sT_{\mathrm{I}}^{2}} \frac{1 + sT_{\mathrm{D}}^{2}}{1 + sT_{\mathrm{D}}'^{2}} = 0. \qquad (54)$$

The following polynomial can be obtained after some manipulations:

$$\begin{aligned} s^6 + &\underbrace{\frac{1}{T_{\mathrm{D}}'^2}}_{p_5} s^5 + \underbrace{\left(-p^2 - \beta K_{\mathrm{PI}}^1 + \alpha K_{\mathrm{PID}}^2 \frac{T_{\mathrm{D}}^2}{T_{\mathrm{D}}'^2}\right)}_{p_4} s^4 \\ &+ \underbrace{\left(-\frac{p^2}{T_{\mathrm{D}}'^2} - \beta K_{\mathrm{PI}}^1 \frac{T_{\mathrm{I}}^1 + T_{\mathrm{D}}'^2}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2} + \alpha K_{\mathrm{PID}}^2 \frac{T_{\mathrm{I}}^2 + T_{\mathrm{D}}^2}{T_{\mathrm{I}}^2 T_{\mathrm{D}}'^2}\right)}_{p_3} s^3 \\ &+ \underbrace{\left(-\gamma K_{\mathrm{PI}}^1 - \beta K_{\mathrm{PI}}^1 \frac{1}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2} + \alpha K_{\mathrm{PID}}^2 \frac{1}{T_{\mathrm{I}}^2 T_{\mathrm{D}}'^2}\right)}_{p_2} s^2 \\ &+ \underbrace{\left(-\gamma K_{\mathrm{PI}}^1 \frac{T_{\mathrm{I}}^1 + T_{\mathrm{D}}'^2}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}\right)}_{p_1} s + \underbrace{\frac{-\gamma K_{\mathrm{PI}}^1}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}}_{p_0}. \end{aligned} \tag{55}$$

From the resulting six equations the following controller parameters can be obtained analytically:

$$T_{\mathrm{D}}'^2 = \frac{1}{p_5}, \quad T_{\mathrm{I}}^1 = \frac{p_1}{p_0} - T_{\mathrm{D}}'^2, \quad K_{\mathrm{PI}}^1 = -\frac{p_0 T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}{\gamma}, \tag{56}$$

the parameter T_{D}^2 is the solution of the second order equation

$$\begin{aligned} \left(p_2 + \gamma K_{\mathrm{PI}}^1 + \frac{\beta K_{\mathrm{PI}}^1}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}\right) \left(T_{\mathrm{D}}^2\right)^2 + \left(-p_3 - \frac{p^2}{T_{\mathrm{D}}'^2} - \beta K_{\mathrm{PI}}^1 \frac{T_{\mathrm{I}}^1 + T_{\mathrm{D}}'^2}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}\right) T_{\mathrm{D}}^2 \\ + \left(p_4 + p^2 + \beta K_{\mathrm{PI}}^1\right) = 0, \end{aligned} \tag{57}$$

and finally

$$T_{\mathrm{I}}^2 = \frac{p_4 + p^2 + \beta K_{\mathrm{PI}}^1}{T_{\mathrm{D}}^2 \left(p_2 + \gamma K_{\mathrm{PI}}^1 + \frac{\beta K_{\mathrm{PI}}^1}{T_{\mathrm{I}}^1 T_{\mathrm{D}}'^2}\right)}, \quad K_{\mathrm{PID}}^2 = \frac{p_4 + p^2 + \beta K_{\mathrm{PI}}^1}{\alpha T_{\mathrm{D}}^2} T_{\mathrm{D}}'^2. \tag{58}$$

The impulse response of the stabilized system can also be seen in Fig. 12. All the closed loop system poles have been set to -1, i.e. the characteristic polynomial is prescribed by $s^6 + 6s^5 + 15s^4 + 20s^3 + 15s^2 + 6s + 1$.

4.3. PD controller for horizontal movement

When $W_{\mathrm{C}_1} = K_{\mathrm{PD}}^1 \frac{1+sT_{\mathrm{D}}^1}{1+sT_{\mathrm{D}}'^1}$, PD-PI and PD-PD controllers result in feasible solution. PD-P controller is redundant and, at the same time, PD-PID is under determined, because there are seven parameters to be determined, but the characteristic polynomial degree is only six.

The PD-PI controller is based on the equation:

$$1 - \frac{\beta s^2 + \gamma}{s^2(s^2 - p^2)} K_{\mathrm{PD}}^1 \frac{1 + sT_{\mathrm{D}}^1}{1 + sT_{\mathrm{D}}'^1} + \frac{\alpha}{s^2 - p^2} K_{\mathrm{PI}}^2 \frac{1 + sT_{\mathrm{I}}^2}{sT_{\mathrm{I}}^2} = 0, \tag{59}$$

from which the following polynomial can be written:

$$\begin{aligned} s^5 + \underbrace{\frac{1}{T_{\mathrm{D}}'^1}}_{p_4} s^4 + \underbrace{\left(-p^2 - \beta K_{\mathrm{PD}}^1 \frac{T_{\mathrm{D}}^1}{T_{\mathrm{D}}'^1} + \alpha K_{\mathrm{PI}}^2\right)}_{p_3} s^3 \\ + \underbrace{\left(-\frac{p^2}{T_{\mathrm{D}}'^1} - \frac{\beta K_{\mathrm{PD}}^1}{T_{\mathrm{D}}'^1} + \alpha K_{\mathrm{PI}}^2 \frac{T_{\mathrm{D}}'^1 + T_{\mathrm{I}}^2}{T_{\mathrm{D}}'^1 T_{\mathrm{I}}^2}\right)}_{p_2} s^2 \\ + \underbrace{\left(\frac{\alpha K_{\mathrm{PI}}^2}{T_{\mathrm{D}}'^1 T_{\mathrm{I}}^2} - \frac{\gamma K_{\mathrm{PD}}^1 T_{\mathrm{D}}^1}{T_{\mathrm{D}}'^1}\right)}_{p_1} s + \underbrace{\left(-\frac{\gamma K_{\mathrm{PD}}^1}{T_{\mathrm{D}}'^1}\right)}_{p_0}. \end{aligned} \tag{60}$$

The analytical solution of the five equations is quite simple:

$$\begin{aligned} T_{\mathrm{D}}'^1 &= \frac{1}{p_4}, \quad K_{\mathrm{PD}}^1 = -\frac{p_0 T_{\mathrm{D}}'^1}{\gamma}, \quad T_{\mathrm{D}}^1 = \frac{p_2 T_{\mathrm{D}}'^1 - p_1 \left(T_{\mathrm{D}}'^1\right)^2 - p_3 + \beta K_{\mathrm{PD}}^1}{\frac{\beta K_{\mathrm{PD}}^1}{T_{\mathrm{D}}'^1} + \gamma K_{\mathrm{PD}}^1 T_{\mathrm{D}}'^1}, \\ T_{\mathrm{I}}^2 &= \frac{p_2 T_{\mathrm{D}}'^1 - p_1 \left(T_{\mathrm{D}}'^1\right)^2 - \gamma K_{\mathrm{PD}}^1 T_{\mathrm{D}}^1 T_{\mathrm{D}}'^1 + p^2 + \beta K_{\mathrm{PD}}^1}{\gamma K_{\mathrm{PD}}^1 T_{\mathrm{D}}^1 + p_1 T_{\mathrm{D}}'^1}, \\ K_{\mathrm{PI}}^2 &= \frac{p_1 T_{\mathrm{D}}'^1 + \gamma K_{\mathrm{PD}}^1 T_{\mathrm{D}}^1}{\alpha} T_{\mathrm{I}}^2. \end{aligned} \tag{61}$$

The PD-PD controller is based on the equation:

$$1-\frac{\beta s^2+\gamma}{s^2(s^2-p^2)}K_{\mathrm{PD}}^1\frac{1+sT_{\mathrm{D}}^1}{1+sT_{\mathrm{D}}^{\prime 1}}+\frac{\alpha}{s^2-p^2}K_{\mathrm{PD}}^2\frac{1+sT_{\mathrm{D}}^2}{1+sT_{\mathrm{D}}^{\prime 2}}=0. \tag{62}$$

The following characteristic polynomial is coming out after some mathematical manipulations:

$$\begin{aligned}
&s^6+s^5\underbrace{\frac{T_{\mathrm{D}}^{\prime 1}+T_{\mathrm{D}}^{\prime 2}}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}}_{p_5}+s^4\underbrace{\left(\frac{1}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}-\beta K_{\mathrm{PD}}^1\frac{T_{\mathrm{D}}^1}{T_{\mathrm{D}}^{\prime 1}}+\alpha K_{\mathrm{PD}}^2\frac{T_{\mathrm{D}}^2}{T_{\mathrm{D}}^{\prime 2}}-p^2\right)}_{p_4}\\
&\quad+s^3\underbrace{\left(-p^2\frac{T_{\mathrm{D}}^{\prime 1}+T_{\mathrm{D}}^{\prime 2}}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}-\beta K_{\mathrm{PD}}^1\frac{T_{\mathrm{D}}^1+T_{\mathrm{D}}^{\prime 2}}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}+\alpha K_{\mathrm{PD}}^2\frac{T_{\mathrm{D}}^{\prime 1}+T_{\mathrm{D}}^2}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}\right)}_{p_3}\\
&\quad+s^2\underbrace{\left(-\frac{p^2}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}-\beta K_{\mathrm{PD}}^1\frac{1}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}-\gamma K_{\mathrm{PD}}^1\frac{T_{\mathrm{D}}^1}{T_{\mathrm{D}}^{\prime 1}}+\alpha K_{\mathrm{PD}}^2\frac{1}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}\right)}_{p_2}\\
&\quad+s\underbrace{\left(-\gamma K_{\mathrm{PD}}^1\frac{T_{\mathrm{D}}^1+T_{\mathrm{D}}^{\prime 2}}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}\right)}_{p_1}+\underbrace{\left(-\frac{\gamma K_{\mathrm{PD}}^1}{T_{\mathrm{D}}^{\prime 1}T_{\mathrm{D}}^{\prime 2}}\right)}_{p_0}.
\end{aligned} \tag{63}$$

The analytical solution of the equations according to this polynomial is tedious. Now, the system of six nonlinear equations has been solved numerically. In this case it is very difficult to find an initial set of the unknown parameters. The experience of this study is that, it is much convenient to use analytically solvable controllers.

4.4. PID controller for horizontal movement

The PID-P controller is redundant, but the other three versions can stabilize the pendulum. The PID-PI controller setup easily can be obtained analytically. The design of a feasible system is based on the characteristic equation

$$1-\frac{\beta s^2+\gamma}{s^2(s^2-p^2)}K_{\mathrm{PID}}^1\frac{1+sT_{\mathrm{I}}^1}{sT_{\mathrm{I}}^1}\frac{1+sT_{\mathrm{D}}^1}{1+sT_{\mathrm{D}}^{\prime 1}}+\frac{\alpha}{s^2-p^2}K_{\mathrm{PI}}^2\frac{1+sT_{\mathrm{I}}^2}{sT_{\mathrm{I}}^2}=0, \tag{64}$$

from which

$$\begin{aligned}
s^6 &+ \underbrace{\frac{1}{T'^1_{\mathrm{D}}}}_{p_5} s^5 + \underbrace{\left(-p^2 - \beta K^1_{\mathrm{PID}} \frac{T^1_{\mathrm{D}}}{T'^1_{\mathrm{D}}} + \alpha K^2_{\mathrm{PI}}\right)}_{p_4} s^4 \\
&+ \underbrace{\left(-\frac{p^2}{T'^1_{\mathrm{D}}} - \beta K^1_{\mathrm{PID}} \frac{T^1_{\mathrm{I}} + T^1_{\mathrm{D}}}{T^1_{\mathrm{I}} T'^1_{\mathrm{D}}} + \alpha K^2_{\mathrm{PI}} \frac{T^2_{\mathrm{I}} + T'^1_{\mathrm{D}}}{T^2_{\mathrm{I}} T'^1_{\mathrm{D}}}\right)}_{p_3} s^3 \\
&+ \underbrace{\left(-\gamma K^1_{\mathrm{PID}} \frac{T^1_{\mathrm{D}}}{T'^1_{\mathrm{D}}} - \beta K^1_{\mathrm{PID}} \frac{1}{T^1_{\mathrm{I}} T'^1_{\mathrm{D}}} + \alpha K^2_{\mathrm{PI}} \frac{1}{T^2_{\mathrm{I}} T'^1_{\mathrm{D}}}\right)}_{p_2} s^2 \\
&+ \underbrace{\left(-\gamma K^1_{\mathrm{PID}} \frac{T^1_{\mathrm{I}} + T^1_{\mathrm{D}}}{T^1_{\mathrm{I}} T'^1_{\mathrm{D}}}\right)}_{p_1} s + \underbrace{\frac{-\gamma K^1_{\mathrm{PID}}}{T^1_{\mathrm{I}} T'^1_{\mathrm{D}}}}_{p_0}
\end{aligned} \tag{65}$$

is the characteristic polynomial, where from the controller parameters can be obtained analytically, $T'^1_{\mathrm{D}} = \frac{1}{p_5}$, furthermore the parameter T^1_{D} is the solution of the second order equation

$$\begin{aligned}
&\left(\frac{\beta p_0}{\gamma T'^1_{\mathrm{D}}} + p_0 T'^1_{\mathrm{D}}\right) \left(T^1_{\mathrm{D}}\right)^2 + \left(-\frac{\beta p_1}{\gamma T'^1_{\mathrm{D}}} - p_1 T'^1_{\mathrm{D}}\right) T^1_{\mathrm{D}} \\
&\qquad + \left(\frac{\beta p_1}{\gamma} - \frac{\beta p_0}{\gamma} T'^1_{\mathrm{D}} + \frac{p_4}{T'^1_{\mathrm{D}}} + p_2 T'^1_{\mathrm{D}} - p_3\right) = 0,
\end{aligned} \tag{66}$$

and finally

$$\begin{aligned}
T^1_{\mathrm{I}} &= \frac{p_1}{p_0} - T^1_{\mathrm{D}}, \quad K^1_{\mathrm{PID}} = -\frac{p_0 T^1_{\mathrm{I}} T'^1_{\mathrm{D}}}{\gamma}, \quad K^2_{\mathrm{PI}} = \frac{p_4 + p^2 + \beta K^1_{\mathrm{PID}} \frac{T^1_{\mathrm{D}}}{T'^1_{\mathrm{D}}}}{\alpha}, \\
T^2_{\mathrm{I}} &= \frac{p_4 + p^2 + \frac{\beta p_0}{\gamma} \left(T^1_{\mathrm{D}}\right)^2 - \frac{\beta p_1}{\gamma} T^1_{\mathrm{D}}}{p_2 T'^1_{\mathrm{D}} - p_1 T^1_{\mathrm{D}} T'^1_{\mathrm{D}} + p_0 T'^1_{\mathrm{D}} \left(T^1_{\mathrm{D}}\right)^2 - \frac{\beta p_0}{\gamma} T'^1_{\mathrm{D}}}.
\end{aligned} \tag{67}$$

The other two controllers, PID-PD and PID-PID, contain seven and eight parameters, respectively, and the characteristic polynomial order is seven. The detailed presentation of these equations are not shown here. The analytical solution of the equations according to these controllers is lengthy and tedious, the system of nonlinear equations can be solved numerically in both cases. It must be highlighted again that, it is very difficult to find a good initial set of the unknown parameters. Anyway it is not a drawback, because there are many other possibilities that can stabilize the pendulum as it is previously shown.

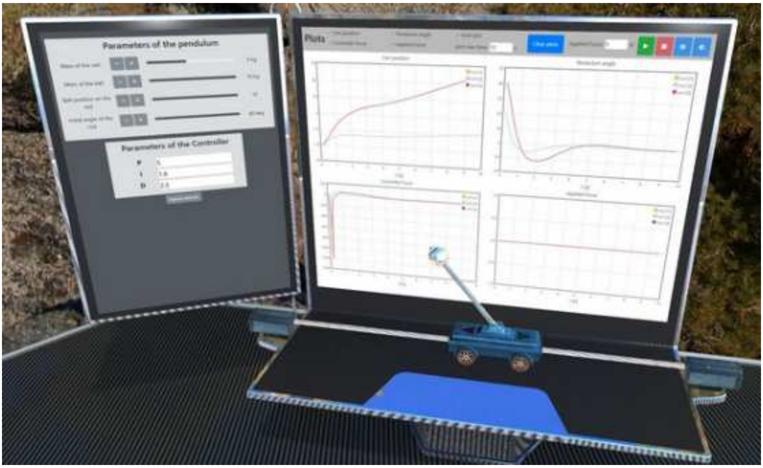


Figure 13. The inverted pendulum model in the virtual laboratory of MaxWhere.

5. Implementation

Controller design and analysis have been realized firstly in Scilab [8]. Then the virtual reality based implementation has been performed in the frame of MaxWhere as a freely available virtual laboratory [9, 10]. A separate future paper is planned to show the virtual lab. A snapshot about the lab can be seen in Fig. 13, where the controller settings and oscilloscopes showing some signals (e.g. the angle of pendulum, the horizontal position, the acting force versus the time) can be seen among other information like the theoretical background.

6. Conclusion and future work

Classical PID controller design for the problem of inverted pendulum has been shown in detail in this paper. Two controllers are necessary to design for stabilizing the system. It is clear that, the design of two dependent controllers is tedious in some cases, however the state feedback controllers can solve this problem in an easy way. Next, the state space representation based controllers will be presented [1–4, 6, 7], like the Ackermann formula and the Bass–Gura pole placement techniques, furthermore the linear quadratic optimal controller design. The Diophantine equation based controller design, the Youla-parametrization [1] are also planned to realize as well as nonlinear techniques [11, 12] and model predictive control [13].

Acknowledgement

Special thank goes to Tamás Budai to include the mentioned techniques in MaxWhere.

References

[1] L. Keviczky, R. Bars, J. Hetthéssy, C. Bányász, Control Engineering, Springer, Singapore, 2019.
doi:10.1007/978-981-10-8297-9_1

[2] B. Lantos, Control Systems, Theory and Design I., Academic Press, Budapest, 2009, in Hungarian.

[3] K. Ogata, Modern Control Engineering, Prentice-Hall, Upper Saddle River, New Jersey, 1997.

[4] P. Gáspár, J. Bokor, A. Soumelidis, An inverted pendulum tool for teaching linear optimal and model based control, Periodica Polytechnica Transportation Engineering 25 (1-2) (1997) pp. 9–19.
URL https://pp.bme.hu/tr/article/view/6592.

[5] V. A. Arya, E. G. Ashni, Stabilisation of cart inverted pendulum using the combination of PD and PID control, International Journal of Innovative Research in Science, Engineering and Technology 7 (4) (2018) pp. 3559–3565.

[6] B. Messner, D. Tilbury, Control tutorials for Matlab Simulink [cited 2019-01-20].
URL `http://ctms.engin.umich.edu/CTMS/`

[7] Md. Akhtaruzzaman, A. A. Shafie, Comparative assessment and result analysis of various control methods, applied on a rotary inverted pendulum, SRV 02 series, Advances in Applied Science Research, Pelagia Research 2 (6) (2011) pp. 83–100.

[8] Scilab [cited 2019-01-20].
URL `https://www.scilab.org/`

[9] Mistems Ltd., MaxWhere [cited 2019-01-20].
URL `https://www.maxwhere.com/`

[10] T. Budai, M. Kuczmann, Development of a VR capable virtual laboratory framework, Pollack Periodica 13 (3) (2018) pp. 83–93.
`doi:10.1556/606.2018.13.3.9`.

[11] B. Lantos, M. Lőrinc, Nonlinear control of vehicles and robots, Springer-Verlag, London, 2011.
`doi:10.1007/978-1-84996-122-6`

[12] A. D. Drexler, Nonlinear and robust control, typescript, Budapest University of Technology and Economics, 2015, in Hungarian.

[13] W. Liuping, Model predictive control system design and implementation using Matlab, Springer-Verlag, London, 2009.
`doi:10.1007/978-1-84882-331-0`