Munich local division
UPC_CFI_180/2025

UPC_CFI_210/2025

Decision

of the Court of First Instance of the Unified Patent Court
Munich local division
issued on 11 March 2026

HEADNOTES

Statements made by a patent applicant during the grant procedure do not constitute interpretative material for the interpretation of the granted patent within the meaning of Article 69(1) EPC. Nor are such statements by the applicant during the grant procedure binding for the interpretation in subsequent infringement or nullity proceedings. However, they may provide an indication of the skilled person's understanding of the teaching of the granted patent at the priority date, because the applicant himself usually has the best understanding of his invention and the relevant technical knowledge (continuation of: Court of Appeal, Order of 20 December 2024, UPC_CoA_402/2024 – Alexion v Samsung Bioepis).

It does not constitute an amendment to the claim within the meaning of Rule 263 of the RoP if, following the defendant's statement of defence, the product in question differs in technical details from the description provided by the claimant in the statement of claim, and the claimant bases the allegation of infringement in the Reply on the technical functioning described by the defendant, provided that the allegation of infringement continues to relate to the generally described product and its contested function.

It is procedurally permissible for the defendant to bring a counterclaim for revocation in response to an infringement action, subject to the condition subsequent that the infringement action is unsuccessful. If this condition is fulfilled because the patent at issue is not infringed, regardless of the validity of the patent, there is no longer any need to rule on the counterclaim for revocation.

HEADNOTES

Statements made by a patent applicant during the grant procedure do not constitute material for the interpretation of the granted patent within the meaning of Article 69(1) EPC, nor are such statements by the applicant in the grant procedure binding for the construction of the claims in subsequent infringement or revocation proceedings. However, they may provide an indication of how a person skilled in the art would understand the technical teaching of the granted patent on the priority date, since the applicant himself usually has the best understanding of his invention and the associated technical knowledge (continuation of CoA, Order of 20 December 2024, 402/2024 – Alexion v Samsung Bioepis).

It does not constitute an amendment to the claim within the meaning of Rule 263 of the Rules of Procedure if, according to the Defendant's Statement of Defence, the product in question differs in technical details from that described by the Claimant in the Statement of Claim and the Claimant bases the allegation of infringement in the Reply to the Statement of Defence on the technical functioning described by the Defendant, provided that the infringement claim continues to relate to the generally described product and its contested function.

It is procedurally admissible for the Defendant to file a counterclaim for revocation in response to an infringement action, subject to the condition subsequent that the infringement action is unsuccessful. If this condition is met because the patent in question is not infringed regardless of its validity, no decision is required on the counterclaim for revocation.

CLAIMANTS, SECOND DEFENDANT AND, AT THE SAME TIME, COUNTER-DEFENDANT

1. **BFexaQC AG**, represented by its Chairman of the Board Bernhard Frohwitter, Südliche Münchner Straße 56, 82031 Grünwald, Germany,

2. **ParTec AG**, represented by its Chairman of the Board Bernhard Frohwitter, Possartstraße 20, 81679 Munich, Germany,

represented by: Dr Roman Sedlmaier, Attorney-at-law, and Jan Gigerich, Patent Attorney, IPCGS Gigerich Sedlmaier Patentanwalt Rechtsanwalt PartGmbB, Paul-Wassermann-Straße 3, 81829 Munich,

supported by: Dr Stefan Richter, Clifford Chance Partnerschaft mbB, Königsallee 59, 40215 Düsseldorf,

Patent attorneys David Molnia and Alexandre Hoffmann, df-mp tech Molnia Ho PartG mbB, Theatinerstraße 16, 80333 Munich, and

Rajvinder Jagdev and Peter FitzPatrick, Irish Solicitors, Powell Gilbert (Europe) LLP, 28-32 Pembroke Street Upper, Dublin, D02 EK84.

DEFENDANTS AND COUNTERCLAIMANTS

1. **NVIDIA Corporation**, represented by its founder, president and CEO, Mr Jensen Huang, 2788 San Tomas Expressway, Santa Clara, CA 95051, USA,

2. **NVIDIA GmbH**, represented by each of the managing directors, Mark Steven Hoose, Ludwig von Reiche, Rebecca Peters, Donald Robertson, Adenauer Straße 20/A4, 52146 Würselen, Germany,

represented by: Johannes Heselberger, Dr Christof Karl, Dr Stefan Lieck, Dr Philipp Bovenkamp, Sabrina Hütt and Henri Kirner, Attorneys-at-law and patent attorneys at Bardehle Pagenberg Partnerschaft mbB, Prinzregentenplatz 7, 81675 Munich.

PATENT AT ISSUE

European Patent No. EP 3 743 812

PANEL/CHAMBER

Panel 2 of the Munich local division

JUDGES

This decision was delivered with the participation of the presiding judge Dr D. Voß (judge-rapporteur), the legally qualified judge Dr G. Werner, the legally qualified judge A. Kupecz and the technically qualified judge A. Dumont.

LANGUAGE OF THE PROCEEDINGS

German

SUBJECT

Action for infringement and counterclaim for annulment

ORAL HEARING

13 February 2026

FACTS

1 The claimants are bringing proceedings against the defendants for infringement of the European patent with unitary effect EP 3 743 812 (UPC_CFI_180/2025). The defendants have brought a counterclaim against the second claimant seeking a declaration of invalidity of the patent at issue (UPC_CFI_210/2025).

2 The patent at issue was filed on 23 January 2019 by the second claimant, claiming the priority of EP 18152903 dated 23 January 2018. The application was published on 2 December 2020. The notice of grant of the patent at issue was published on 2 August 2023. Registration as a patent with unitary effect, applied for on 30 August 2023, took place on 6 September 2023 (see Annex K 45). The contracting member states at the time of registration of the unitary effect were Austria, Belgium, Bulgaria, Germany, Denmark, Estonia, Finland, France, Italy, Lithuania, Luxembourg, Latvia, Malta, the Netherlands, Portugal, Sweden and Slovenia.

3 The patent at issue, the language of the proceedings for which is English, relates to the dynamic allocation of heterogeneous computing resources determined over the course of the application. Claim 1 of the patent at issue protects a method and reads as follows:

> “A method of operating a heterogeneous computing system (10) comprising a plurality of computation nodes (20) and a plurality of booster nodes (22), at least one of the plurality of computation nodes (20) and the plurality of booster nodes (22) being arranged to compute a computation task, the computation task comprising a plurality of sub-tasks, wherein

in a first computing iteration, the plurality of sub-tasks are assigned to and processed by ones of the plurality of computation nodes (20) and booster nodes (22) in a first distribution;

characterised in that

information relating to the processing of the plurality of sub-tasks by the plurality of computation nodes (20) and booster nodes (22) is used to generate a further distribution of the sub-tasks between the computation nodes (20) and booster nodes (22) for processing thereby in a further computing iteration.

4 In the German translation, claim 1 of the patent at issue reads:

"A method for operating a heterogeneous computing system (10) comprising a plurality of computing nodes (20) and a plurality of booster nodes (22), wherein at least one of the plurality of computing nodes (20) and the plurality of booster nodes (22) is configured to compute a computational task, wherein the computational task comprises a plurality of subtasks,

wherein, in a first iteration of the computation, the plurality of subtasks are allocated in a first distribution to one of the plurality of compute nodes (20) and booster nodes (22) and are processed by them;

characterised in that

information relating to the processing of the plurality of subtasks by the plurality of computing nodes (20) and booster nodes (22) is used to generate a further distribution of the subtasks amongst the computing nodes
(20) and booster nodes (22) to process them in a further computational iteration."

5 The second claimant is the sole owner of the patent at issue, as registered in the patent register.

6 On 29 August 2021, Mr Bernhard Frohwitter, Chairman of the Board of Directors of both Claimants, signed an agreement on their behalf, whereby the first Claimant was to be granted an exclusive licence to the patent at issue for the field of microelectronics, including the development, manufacture and distribution of microchips and microprocessors. For further details, reference is made to Annex K 50.

7 A further agreement, intended to grant FL Systems AG & Co. KG an exclusive licence, inter alia, to the patent at issue for the field of application of supercomputer system architecture (high-performance computing), was signed on 21 December 2023 by Mr Frohwitter on behalf of the second claimant and FL Systems AG & Co. KG. The second claimant is the general partner of FL Systems AG & Co. KG. On 7 June 2024, Mr Frohwitter also signed, on behalf of FL Systems AG & Co. KG and the

Claimant 1) a contract whereby the former sub-licensed the exclusive licence granted to it. For the details of the two licence agreements, reference is made to Annexes K 48 and K 49.

8 The first defendant develops graphics processing units. It manufactures Nvidia DGX systems, which it promotes, amongst other things, on its German-language website "https://www.nvidia.com/de-de/datacenter/dgx-systems/" with reference to the products "Nvidia DGX A100", "Nvidia DGX H100", "Nvidia DGX BasePOD" or "Nvidia DGX SuperPOD" (Exhibit K 8). In the context of presenting its DGX systems, Defendant 1) points out that Nvidia DGX systems can be purchased through certified partners, including, for the Federal Republic of Germany, the “NPN Elite Solution Providers” “Amber”, “Delta” and “sysGen” (Exhibits K 9 and K 10). Several hundred solution providers from a wide range of countries are listed in the complete list of NPN partners (Exhibit K 11). A blog post dated 1 May 2023 on the website of Defendant 1) also promotes the global use of DGX H100 systems, including for DeepL in the Federal Republic of Germany (Exhibit K 12).

9 The second defendant is part of the Nvidia Group. It is controlled by Nvidia Limited, which has its registered office in London and is itself included in the consolidated financial statements of the first defendant. According to the extract from the commercial register, the business of the second defendant comprises the development of semiconductors and other computer hardware and software components, and the distribution of its own and third-party hardware and software products (Exhibit K 6). The second defendant
is also active in the fields of research and development, as well as sales promotion and marketing (Exhibit K 4). It also assumes responsibility for ensuring product reliability and quality management on behalf of manufacturers, assembly firms and distributors who use Nvidia technology in their products or sell Nvidia products. The DGX H100 model also bears the CE marking, in the obtaining of which the second defendant plays an active role. It is the only European company from which a copy of the declaration of conformity can be requested directly.

10 The claimants’ action is directed against the defendant’s DGX system products.

11 Nvidia DGX is a range of Nvidia servers and workstations which, in addition to using CPUs, specialise primarily in the use of graphics processing units (GPUs) to accelerate deep learning applications. The DGX system "DGX H100" forms the basic building block for large AI clusters such as DGX POD or DGX SuperPOD and is thus designed as the centrepiece of an AI competence centre for businesses. Other DGX systems such as the DGX A100, DGX H200 or DGX GH200 use different components, but do not differ conceptually from the DGX H100 model. All DGX servers and workstations are identical in terms of the alleged infringement. They contain GPUs combined with CPUs featuring a large number of CPU cores. When workloads are allocated to these products, the CPUs act as a central

management centre for the GPUs, distributing the tasks and collecting the available results.

12 The Run:ai software from Run:ai can be installed on Nvidia DGX systems. The Run:ai Cluster Engine handles the dynamic management of AI workloads and resources, with a particular focus on optimising GPU utilisation. It consists of two components: on the one hand, the Run:ai clusters for scheduling services and workload management, and on the other hand, the Run:ai control plane for resource management, workload submission and cluster monitoring. At the turn of the year 2024/2025, the Nvidia Group acquired Run:ai.

APPLICATIONS

13 In their action for infringement, the claimants seek:

I. The defendants be ordered to refrain from

devices suitable for carrying out a method for operating a heterogeneous computing system (10) comprising a plurality of computing nodes (20) and a plurality of booster nodes (22), wherein at least one of the plurality of compute nodes (20) and the plurality of booster nodes (22) is configured to compute a computational task, wherein the computational task comprises a plurality of subtasks,

to customers within the territory of the Republic of Austria, the Kingdom of Belgium, the Republic of Bulgaria, the Kingdom of Denmark, the Republic of Estonia, the Republic of Finland, the French Republic, the Federal Republic of Germany, the Italian Republic, the Republic of Latvia, the Republic of Lithuania, the Grand Duchy of Luxembourg, the Republic of Malta, the Kingdom of the Netherlands, the Portuguese Republic, the Republic of Slovenia and/or the Kingdom of Sweden,

wherein the method comprises at least:

in a first iteration of the calculation, the plurality of subtasks are allocated in a first distribution to one of the plurality of compute nodes (20) and booster nodes (22) and processed by these; characterised in that

Information relating to the processing of the plurality of subtasks by the plurality of compute nodes (20) and booster nodes (22) is used to generate a further distribution of the subtasks amongst the compute nodes (20) and booster nodes (22) for processing in a further computational iteration,

– indirect infringement of claim 1 of EP 3 743 812 B1 –

in particular, where in the method

each of the plurality of subtasks may be more or less suitable for processing by one or more of the computing nodes or one or more of the booster nodes;

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 1 –

and/or, where in the method

a resource manager determines the allocation of tasks and subtasks to the compute nodes and booster nodes for the first iteration, depending on the computational task;

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 2 –

and/or, if in the method

the initial distribution of subtasks between compute nodes and booster nodes is based on the scalability of the subtask;

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 3 –

and/or, if in the method

a resource manager dynamically changes the mapping of compute nodes and booster nodes to one another during the computation of the computational task on the basis of the information;

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 4 –

and/or, if in the method

an application manager receives the information and determines the further distribution, and wherein a resource manager determines the allocation of tasks and subtasks to the compute nodes and booster nodes for the first iteration depending on the computational task, and wherein the application manager receives the information and processes it as input for the resource manager, so that the resource manager dynamically changes the further distribution during the computation of the computational task.

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 5 –

and/or, where in the method

an application manager receives the information and determines the further distribution, and wherein a resource manager determines the allocation of tasks and subtasks to the compute nodes and booster nodes for the first iteration depending on the computational task, and wherein the application manager receives the information and processes it as input for the resource manager, such that the resource manager dynamically modifies the further distribution during the computation of the computational task, and wherein the resource manager receives the information, such that the resource manager dynamically modifies the mapping of the compute nodes and booster nodes to one another during the computation of the computational task.

– indirect infringement of claim 1 of EP 3 743 812 B1
as amended by auxiliary claim 6 –

II. The defendants are further ordered, within a period of 30 days of service of the notice within the meaning of Rule 118(8) sentence 1 of the RoP and, where applicable, the certified translation,

to provide the claimants with information on the extent to which they have committed the acts described in Section I since 2 January 2021, in the form of a breakdown of the following information, structured for each month of a calendar year and according to the products described in Section I:

a) the origin and distribution channels of the infringing products;

b) the quantities produced, manufactured, delivered, received or ordered, and the prices paid for the infringing products;

c) the identity of all third parties involved in the manufacture or distribution of the infringing goods;

d) the number and details of the products offered;

e) the advertising carried out, broken down by advertising medium, its reach, the period of distribution and the distribution area; including evidence of these advertising activities;

f) the costs, broken down by individual cost factors, and the profits made,

whereby, to substantiate the information, copies of the relevant purchase documents (namely invoices, or alternatively delivery notes) must be submitted, with details requiring confidentiality being redacted from the data subject to disclosure and notification;

III. The defendants are further ordered

1. in the event of any breach of the order pursuant to the application under point I, to pay a repeated penalty payment of at least EUR 1,000.00 per infringing product;

2. in the event of any breach of the order pursuant to the application under point II, to pay a recurring penalty of at least EUR 250.00 per day to the court for each day of the breach.

IV. The defendants are ordered to pay the claimants EUR 100,000.00 as provisional damages, which shall be adjusted if the acts referred to in paragraph I continue.

V. It is hereby declared

1. that the defendants are obliged to pay the claimants appropriate compensation for the acts described in I. and committed in the period from 2 January 2021 to 1 September 2023;

2. that the defendants are obliged to compensate the claimants for all damage suffered by them as a result of the acts referred to in I. committed since 2 September 2023 and which will be suffered in the future.

VI. The defendants are further ordered to bear the costs of the proceedings.

14 The defendants request that

I. The claim is dismissed. In

the alternative:

Ia. The payment of provisional damages pursuant to Rule 119 of the RoP is conditional upon the Claimants providing security (by way of a cash deposit or bank guarantee) to the defendants in an amount equal to the provisional damages awarded (Rule 352 of the RoP).

II. The claimants shall bear the costs of the proceedings jointly and severally.

15 In their reply of 13 June 2025, the claimants also asserted for the first time a direct infringement of the patent at issue based on patent claim 1 and announced corresponding applications. The admission of this extension of the claim was rejected by order of 15 January 2026.

16 The defendants have brought a counterclaim for revocation. When asked at the oral hearing whether the defendants would wish a decision on the counterclaim if the court were to conclude elsewhere that the

patent at issue is not infringed, the defendants stated that they agreed to such a condition for the counterclaim for revocation. They therefore request:

I. Patent EP 3 743 812 is revoked to the extent of claim 1 as granted and to the extent of claim 1 of the respective auxiliary request.

II. The Claimants submit the costs of the counterclaim proceedings jointly and severally.

17 The claimants have stated that, in view of other potential infringers, they have an interest in a decision on the counterclaim and request that

1. the defendant's counterclaim seeking the revocation of European Patent EP 3 743 812 be dismissed;

   in the alternative

   to uphold European Patent EP 3 743 812 in the version set out in one of the alternative claims 1 to 6 in accordance with Annex K 60 (alternatively in that order); and

2. order the defendants to pay the costs of the counterclaim.

18 In the context of the counterclaim for revocation, the defendants, following the expiry of the time limits for filing pleadings, submitted the document "Efficient CPU-GPU cooperative computing for solving the subset-sum problem" by Wan et al. (Annex BP CR 11) as new prior art by document dated 24 January 2026. On 30 January 2026, the judge-rapporteur granted the requested admission of the document, setting a deadline for the claimants to submit their comments on its content. He reserved the decision on the admission of the additional prior art and the associated extension of the claim to the panel.

19 In this regard, the defendants request that

to admit the additional prior art (document by Wan et al.) into the proceedings.

20 The claimants request that

that the extension of the claim not be permitted.

POINTS OF DISPUTE BETWEEN THE PARTIES

Standing to sue and entitlement to claims

21 The claimants are of the view that the licence agreements concluded between them and FL Systems AG & Co. KG are valid. The defendants have not demonstrated that a sham transaction took place. There is no prohibited self-dealing transaction, as Mr Frohwitter is exempt from the prohibition on multiple representation for all companies, as evidenced by the extracts from the commercial register submitted. The licence agreements cover the full scope of protection and application of the patent at issue, meaning that Claimant 1 is licensed for the entire subject-matter of the patent at issue. Insofar as this is necessary at all for the standing to sue or any claims of the second claimant, the second claimant, as general partner of FL Systems AG & Co. KG, also participates economically in the granting of the licence.

22 The defendants contend that the two licence agreements concluded by the second claimant with the first claimant and FL Systems AG & Co. KG are invalid. They argue that these constitute a sham transaction because no consideration was agreed. The defendants have recently dropped the objection of multiple representation raised against all the licence agreements.

23 The defendants are of the view that, in view of the licences granted, Plaintiff 2 lacks standing to sue, or at any rate lacks entitlement to the claim. For the latter, it must be demonstrated that she is the holder of the claims she is asserting. Where there are several parties with standing to sue who are asserting the same claim, all claimants must be the holders of that claim for the period in question. However, this is not the case here. Claimant 2) has relinquished all rights of use through the granting of licences.

24 Claims for damages and related claims for information covering the period up to 1 September 2023 were already ruled out simply because no such claims had been raised and the necessary conditions were not met. For instance, the translation of the patent specification into the respective national language, as required under French and German law, did not exist. Furthermore, an indirect infringement does not give rise to any claims for damages in any event. Further claims for injunctive relief, damages and information were also ruled out on account of the exclusive licences, but at the latest with the licence granted to FL Systems AG & Co. KG from 21 December 2023. The second claimant was also not affected by a patent infringement, as it had not reserved any rights of use in the contracts. The defendants dispute that the second claimant has any economic interest in the granting of the licence. In their view, this does not necessarily follow from the second claimant's position as general partner of FL Systems AG & Co. KG. Typically, only the limited partners would share in the profits of a limited partnership.

25 Claimant 1) is entitled to claim at the earliest from 7 June 2024 and the conclusion of the licence agreement with FL Systems AG & Co. KG.

Interpretation

26 The claimants take the view that the computing and booster nodes referred to in the patent in suit need not necessarily be autonomous or independent computing units. If one were to refer to WO 2012/049247, mentioned in the patent at issue, for the interpretation of the patent claim, then it must be taken into account that it discloses both individual processors as nodes and, furthermore, that the order of computing and booster nodes therein is designed to enable virtualisation at all relevant levels. Consequently, a ‘computing node’ could simply be a multi-core processor and a ‘booster node’ a graphics processing unit (GPU), a many-core processor or a cluster of many-core processors.

27 Computing and booster nodes are therefore designed differently and have different technical capabilities. Functionally, whilst computing nodes and booster nodes are different, they are complementary in their ability to compute heterogeneous tasks comprising various types of tasks or subtasks. The processing of the computational task is achieved through an appropriate distribution of tasks across computing and booster nodes.

28 The claimants further contend that the concept of a computational task with its subtasks cannot be equated with a (homogeneous) data set. A computational task and its subtasks comprise instructions or commands which, unlike simple data sets, are not readily divisible and are consequently typically heterogeneous.

29 As regards the distribution of the sub-tasks, in the claimants’ view, claim 1 of the patent at issue presupposes a direct distribution, if only to be able to obtain qualified information. This would not be possible in the case of an indirect distribution in which the computing nodes were involved.

30 In so far as claim 1 of the patent at issue requires that a further distribution of the sub-tasks take place in order to process them in a further iteration, the claimants argue that these need not be exactly the same sub-tasks that were the subject of the first distribution. Rather, the focus is on the subtasks that are the subject of the further iteration. It need not necessarily be the same set of subtasks as in the first iteration. The number of subtasks may vary. Subtasks that are the subject of the second distribution may also be newly defined and need not have already been the subject of the first distribution. Furthermore, the ‘further computational iteration’ does not necessarily constitute the second iteration following the first iteration. Finally, in the further computational iteration, it is also not necessary for there to be a different distribution of sub-tasks across all sub-tasks than in the first iteration. It is also possible to distribute only some of the sub-tasks to other nodes.

31 In the claimants' view, the concept of 'computational iteration' must be distinguished from a mere iteration and encompasses a processing step or a processing period following which an assessment of the system's status and the initial distribution can reasonably be carried out, and meaningful information is available for further, essentially system-wide processing or distribution. Claim 1 of the patent at issue does not require that the processing of the subtasks be completed in the first iteration. Computational tasks for high-performance cluster computer systems are generally designed for processing over days, weeks and months. They may also be paused or suspended during this time, but do not necessarily have to be completed. However, this is not ruled out either, meaning that subtasks could then be repeated in a further iteration with different results.

32 As regards the distribution of subtasks, the initial distribution could be based on the computational task – such as its scalability – but was not limited to this. It could also be based on other factors such as the status and availability of resources, the interdependence of subtasks, or similar considerations.

33 In the claimants' view, the information on the basis of which further distribution is to take place according to the teaching of the patent at issue could relate to the processing of the subtasks, or, where applicable, to scalability or status information regarding the computing and booster nodes. The fact that the latter was already known in the prior art for load balancing is irrelevant. The patent at issue mentions both types of information and differs from the prior art ('load balancing') in that the prior art did not recognise that the processing of the subtasks has a further-reaching influence on the status of the nodes and does not merely concern their utilisation, for example when considering the suitability of the node for a computational task. The patent at issue differs from this by using qualified information, because information regarding the mere utilisation of the nodes is too coarse-grained.

34 The defendants are of the view that the terms 'computing node' and 'booster node' in claim 1 of the patent at issue should be defined as set out in the prior art WO 2012/049247, on which the subject-matter of the patent at issue is expressly based. According to the granted version EP 2 628 080 of this prior art, a node in a cluster is an autonomous/independent computing unit, for example in the form of a 'stand-alone computer'. This is also apparent directly from the WO publication. Furthermore, the booster nodes are separate units that are not part of the computing nodes. In so far as the claimants consider that computing and booster nodes are suitably suited to different subtasks, the defendants take the view that this is possible, but that claim 1 of the patent at issue is not limited to this. Nor are there any apparent functional reasons for different characteristics.

35 As regards the (sub-)tasks, these need not be ‘heterogeneous (sub-)tasks’, and in particular not (sub-)tasks which could be distinguished on the basis of whether they are better suited to being processed on computing nodes or on booster nodes. This has not been received by claim 1; rather, the description of the patent in suit suggests otherwise.

36 With regard to the allocation of subtasks, the defendants take the view that this does not necessarily have to take place at the distribution level, and therefore directly, and in particular not by a central unit. The decision on allocation to individual nodes could also be made by computing nodes, which (further) distribute subtasks to booster nodes connected to them.

37 Claim 1 of the patent at issue requires, following the first distribution of the plurality of subtasks, a second distribution that differs from the first. However, it follows from the wording of the claim that all subtasks comprising the computational task must form part of both the first and second iterations. The same set of subtasks must be processed in the first and second iterations. The increase in efficiency associated with the redistribution only occurs when the same subtasks are used in the multiple iterations.

38 Furthermore, it follows from the concept of ‘iteration’ that each of the subtasks must be calculated at least twice, namely in a first and a second iteration. This is because the term denotes the repeated calculation of the same set of subtasks.

39 The information referred to in claim 1, which relates to the processing of the plurality of subtasks by the plurality of computing and booster nodes, must be distinguished from information concerning the utilisation or status of the computing nodes and booster nodes. Only such information which – like the scalability of subtasks – relates to the actual, initial processing of the subtasks is claimed, since, according to the invention, the system is intended to learn from this information for the second/further iteration, i.e. the reprocessing of the same subtasks. The use of the nodes’ status information for further distribution is not according to the invention, but is prior art. Furthermore, the information relied upon by the claimants regarding resource availability or the utilisation/capacity of nodes constituted nothing more than status information, which is not within the scope of the claims.

Contested embodiment

40 The claimants stated in the statement of claim that the embodiment challenged in the action consists of DGX system products supplied with the Run:ai software. In response to the defendant’s contention that the software was not pre-installed, the claimants maintained in their Reply their assertion that DGX system

products are offered and supplied with Run:ai software. At the same time, they stated that even if Run:ai were not offered and supplied as part of a ‘bundle’ with the contested DGX systems, but had to be purchased and installed separately, there could be no doubt that an indirect infringement had taken place.

41 In this regard, the claimants claim that the Nvidia DGX systems are supplied as standard with software from the company Run:ai. The Run:ai Documentation Library states in this regard: “NVIDIA DGX comes bundled out of the box with Run:ai.” (Exhibit K 21). The software has been fully tested and certified for use on clusters of DGX systems. Even if Run:ai has to be installed by the end user on the purchased DGX systems, Nvidia provides detailed instructions for this and even integrates the installation of Run:ai into its own Base Command Manager software (“BCM”), which is installed on all DGX systems (Exhibit K 52).

42 Ultimately, in the claimants’ view, it is irrelevant whether the customer has to install the Run:ai software themselves, because in a complex high-performance computing system used for applications such as machine learning, the installation of a whole range of different software programmes is essential. Any requirement for the separate installation of Run:ai on the DGX systems is irrelevant to the question of whether the defendants market DGX systems together with Run:ai as a ‘bundle’ or ‘package’. In 2023, DGX BasePod devices were in fact offered with the Run:ai software. Nvidia also provides first-level support for Run:ai and has integrated Run:ai into the software stack of the DGX Cloud Create service. In light of all this, the Run:ai software – regardless of the method of distribution or installation – is an essential component of Nvidia’s bundle and therefore an integral part of its offering.

43 The defendants claim that the Run:ai software is not yet installed when a DGX product is purchased. The Run:ai software must be downloaded and installed independently by the defendants’ customers or by someone acting on their behalf. This also applies to the period following Nvidia’s acquisition of Run:ai. Nor does anything to the contrary arise from the software documentation (Exhibit K 21). It also states that Run:ai must be installed and set up on the DGX hardware only after its purchase. The description ‘bundled out of the box with Run:ai’ simply means that the DGX products are designed in such a way that the Run:ai software can be installed and used on them by the defendants’ customers, but is not pre-installed. This is because the Run:ai software is not required for the DGX systems to function properly. In order to install the Run:ai software, the purchaser of the DGX system must purchase a separate software subscription

44 The defendants contend that the claim is directed exclusively against the defendants’ DGX hardware products, which are allegedly supplied together with the Run:ai software

. No such embodiments exist. Insofar as the claimants are now attempting to broaden the scope of the contested embodiment, such a broadening is inadmissible, as it must be assessed against the requirements of Rule 263 of the RoP. There is no corresponding application to that effect.

Infringement

45 The claimants are of the opinion that the contested embodiment is suitable for applying the protected method, regardless of whether the Run:ai software is pre-installed or not. This can be demonstrated, for example, using the DGX H100 system.

46 The contested embodiment constitutes a heterogeneous computing system within the meaning of the patent at issue. In the case of the DGX H100 system, it comprises a plurality of compute nodes in the form of the Intel Xeon 8480C CPUs used and a plurality of booster nodes in the form of the NVIDIA H100 Tensor Core GPUs. Both individual DGX servers (Single Appliance Read) and several of these servers interconnected via a high-performance network to form a DGX SuperPOD system (Multi Appliance Read) each constitute a heterogeneous computing system within the meaning of the patent at issue, their processors being the respective ‘nodes’.

47 In the defendant’s view, in accordance with the functionality of the Run:ai software, computing tasks – referred to therein as ‘workloads’ – could consist of several sub-tasks, also known as ‘pods’, with each pod being executed on its own node – CPU or GPU. The Run:ai scheduler makes decisions regarding resource allocation on the basis of predefined Rules. When a workload is submitted for processing, the Run:ai scheduler checks whether sufficient resources are available for this task by comparing the request with the currently allocated resources and the specified quotas. Consequently, in an initial iteration of the calculation, the multitude of subtasks are allocated in an initial distribution to one of the multitude of compute and booster nodes and processed by these.

48 However, the scheduler constantly recalculates the fair allocation of resources for each project and redistributes the resources accordingly (Appendix K 40). The Run:ai software is able to dynamically scale workloads up or down whilst the system is running, a process known as elasticity. If additional resources are available, the elasticity function allows these to be added to the workloads during runtime to accelerate the execution of the workload. Consequently, information relating to the processing of the subtasks is used to generate a further distribution of the subtasks among the nodes for processing in a further computational iteration.

49 The Run:ai scheduler allocates resources to the workload not only according to their availability, but also in accordance with Run:ai’s quota and fairness principles. This results in

that, in certain scenarios, the Run:ai Scheduler reallocates resources from one workload to another in accordance with these fairness principles. If the Scheduler has made an allocation of resources that exceeds the requirements for a workload in a scheduling cycle, it may make a subsequent allocation of resources to another workload that requires resources. In the claimants' view, the resources actually allocated to a workload in a previous planning cycle constitute information relating to the processing of the subtasks. This is because the status of the nodes and, more generally, how a distribution could be better organised based on the situation in the DGX device also constitute information relating to the calculation of the subtasks within the meaning of the patent at issue, as they could influence the further distribution of the tasks among the nodes.

50 The claimants are of the view that the consideration of information in accordance with the patent for the further distribution of subtasks for a further computational iteration also arises from a function of the Run:ai software that supports 'Hugging Face' inference models. Accordingly, every time a variable condition relating to the underlying inference workload (i.e. the set of subtasks) is met, Run:ai creates a new replica pod (i.e. a new subtask) and performs resource allocation for it. The relevant variables – latency, throughput and parallelism – represent information relating to the processing of the subtasks from which the inference workload is composed.

51 In the claimants' view, each planning cycle of the Run:ai scheduler constituted an iteration within the meaning of the patent at issue. There were several such iterations. The calculation of the workload could be interrupted and the resources allocated to it could be changed in the meantime. These 'interruptions' defined the planning cycle of the Run:ai scheduler. This applied to both the Single Appliance Read and the Multi-Appliance Read. It was not necessary for the subtasks to be completed within each iteration.

52 The defendants contend that the contested embodiment does not constitute a heterogeneous computing system as claimed, comprising a plurality of computing and booster nodes. Rather, it is a single server, that is to say a single
'node' containing CPUs and GPUs, but not independent computing or booster nodes as required by claim 1. CPUs and GPUs are not independent computing units within the meaning of the patent at issue; in particular, the GPUs are not booster nodes, but components of a DGX system and thus of a computing node. According to the Run:ai documentation, nodes are independent servers comprising various components such as CPUs and GPUs, which are not themselves nodes.

53 The defendants share the view that the 'workloads' referred to by the Run:ai software could be regarded as computational tasks and the 'pods' as subtasks. The

defendants claim, however, that [redacted]

54 Insofar as the scheduler of the Run:ai software ensures dynamic resource allocation, whereby the ‘elasticity feature’ enables the use of additional resources – where available – for task processing, the teaching of the patent at issue is precisely not realised. The allocation of workloads to the CPUs and GPUs of the contested embodiment is based on the availability of the resource, i.e. the status or utilisation of the resources. This availability does not constitute information within the meaning of the patent at issue that relates to the processing of the subtasks.

55 In the defendant’s view, the ‘priorities’ referred to by the claimants do not constitute information relevant to the claim that relates to the processing of the workload’s pods. They are set by the users and, moreover, merely determine the prioritisation of workloads relative to one another; they play no role in the processing of an individual workload, and even less so in the processing of the pods.

56 Similarly, the so-called ‘fair share’ does not constitute information relevant to the claim relating to the processing of the pods of a workload. The fair share merely indicates the sum of the guaranteed resources plus the proportion of non-guaranteed resources for a project. The fair share therefore also relates solely to information regarding the availability of resources. The information regarding resources actually allocated to a workload in a previous scheduling cycle does not constitute information relating to the processing of subtasks, but rather information concerning the availability and utilisation of the GPU. The same applies to information regarding resources to which no pods are currently assigned.

57 As regards the “Hugging Face inference workload”, the defendants are of the view that this argument could also have been raised in the infringement action. Regardless of this, a “Hugging Face inference workload” is not a computational task as defined in the claim. Rather, the pods of a “Hugging Face inference workload” – just like normal workloads – are only computed once. There is no further iteration. Furthermore, there is no second distribution. The ‘Hugging-Face inference workloads’ are based on the principle that one or more new ‘replicas’ are generated as required. However, these replicas are not sub-tasks as claimed, but merely a copy of the first (part of the) workload. If one assumes a first and second iteration of replicas, the same pods would not be redistributed. Rather, these would then be iterations of different, and no longer identical, pods, because the replicas had been newly added and different pods were now being processed. The mere addition of replicas also does not constitute redistribution, because the distribution of the remaining pods does not change.

58 [redacted]. There is no interruption in the calculation of the pods following a complete calculation/iteration of the pods. Furthermore, the calculation of the pods is carried out solely on the basis of a fixed scheduling cycle. This is independent of the calculation of the respective workloads or pods, but runs globally.

59 In the absence of further iterations, the defendants argue, there is also no change in the allocation,
i.e. the redistribution of pods takes place between two iterations.

60 Finally, the defendants take the view that a second distribution of the subtasks (pods) of a computational task (workload) does not take place simply because the subtasks of the first distribution are not the same subtasks that are the subject of the alleged second distribution. Pods of a workload to which resources had been allocated in a first planning cycle, but which had been allocated to other workloads in a subsequent planning cycle, would not be redistributed; rather, their processing would be interrupted and deferred. Thus, the 'same' sub-tasks would not be processed before and after the second planning cycle.

Acts of infringement and legal consequences

61 The claimants are of the view that the defendants are liable for indirect patent infringement, regardless of whether the Run:ai software is pre-installed. The DGX systems, in and of themselves and without Run:ai, already embody numerous claim features and are thus "means relating to an essential element of the invention." Furthermore, the preparation for and coordination with Run:ai demonstrates that the DGX systems are suitable and intended for use in accordance with the patent at issue, in particular with Run:ai. The defendants were also aware that the DGX systems were intended and suitable for the application of the patent-protected method and that the purchaser of the systems would also use them in this manner within the scope of the patent at issue.

62 The defendants argue that there is no indirect patent infringement because the contested embodiment – DGX System products with pre-installed Run:ai software – does not even exist. There is no contested means within the meaning of Article 36 of the UPC Agreement. The actions of Defendant 1 do not constitute an indirect infringement of claim 1 of the patent at issue. Defendant 1 merely manufactures the DGX systems. However, the software is installed by the customers, who then also apply the protected method. The same applies to Defendant 2) as to Defendant 1). All alleged acts of infringement by Defendant 2) relate exclusively to the DGX products manufactured by Defendant 1). Furthermore, the 'sales promotion campaigns' and claimed by the Claimants, and Defendant 2)'s 'warranty for

product reliability and quality management" of the second defendant with regard to the DGX systems are not described in further detail, let alone substantiated by offers of evidence.

63 With regard to any legal consequences, the defendants argue that no fixed deadline should be set for the provision of information. Nor should all the requested details be provided as part of the information. An order to destroy, recall and/or remove the products from distribution channels would be disproportionate. The imposition of penalty payments should not be made dependent on fixed amounts and deadlines in advance. Furthermore, the amount and the conditions for adjusting the provisional damages are unclear. In any event, the claimants would have to provide security in advance in this regard. The application for a declaration of liability for damages is indefinite; in any case, there is no class of creditors.

Counterclaim for annulment

64 The defendants consider claim 1 of the patent at issue to be unpatentable.

65 The technical teaching of the patent claim is prejudicially anticipated by patent application US 2012/0233486 A1 ('Phull'). In particular, Phull discloses the allocation of subtasks to booster nodes. In this respect, it is irrelevant if the booster node does not directly participate in the distribution of the subtasks, but is allocated to a CPU which delegates tasks to it (so-called 'offloading'). Notwithstanding this, Phull also discloses standalone GPUs. Insofar as Phull describes that data sets are redistributed and allocated, the processing of data also constitutes a computational task within the meaning of the patent at issue.

66 In the defendant's view, the technical teaching of the patent at issue is also not novel in relation to US 2017/0109207 A1 ('Li'). In this respect, it is sufficient for computing nodes and booster nodes to have different computing power, even if they are otherwise structurally of the same nature. Information relating to the processing of the subtasks within the meaning of the patent at issue is also used. A query as to the extent to which a node has processed a volume of data provides such information, which consequently relates to the processing of the multitude of subtasks by the multitude of computing nodes and booster nodes.

67 In any event, however, the teaching of the patent at issue is suggested by 'Li' in combination with the prior art or 'Phull'.

68 The defendants further contend that the additional prior art submitted with the rejoinder of 13 August 2025 should be admitted into the proceedings. In support of this, they argue that the additional prior art was submitted at the earliest possible opportunity because, in their Reply, the claimants had relied on an interpretation of the claims that differed from their earlier interpretation and from the interpretation made during the grant proceedings, and was therefore not to be expected in this form. According to the new

interpretation, the claimants now wish to understand information regarding the status (i.e. the availability) of nodes in a heterogeneous computing system to be included under information relating to the calculation of sub-tasks. On the basis of this interpretation, the subject-matter of the patent at issue is not novel in view of Deng's article "Method for scheduling data flow tasks and apparatus", but is in any event suggested by a combination of WO 2912/049247 A1 or Deng with Zaki's publication "Customized Dynamic Load Balancing for a Network of Workstations".

69 Furthermore, the defendants contend that the Wan citation should be admitted into the proceedings. This was not found despite two search requests to two different reputable search firms. It was only through a search conducted for another client of the law firm representing the defendants that they became aware of this citation on 14 January 2026. It is in the nature of things that the results of a search for relevant prior art cannot claim to be exhaustive. This applies all the more to scientific publications from a non-top-tier journal.

70 In the defendant's view, Wan anticipates the teaching of the patent at issue in a manner detrimental to novelty. In particular, according to the claimants' own interpretation, the sub-tasks processed in the second iteration need not be identical to those of the first iteration. On this assumption, the teaching of the patent at issue is disclosed in Wan.

71 The claimants consider claim 1 of the patent at issue to be patentable. The invention according to the patent is novel in relation to 'Phull'. The prior art discloses neither a heterogeneous computational task divided into sub-tasks, nor a distribution of sub-tasks to computational and booster nodes interacting in accordance with the patent, as required by the claimed technical teaching.

72 In the claimants' view, Li is also not detrimental to novelty. The prior art does not disclose booster nodes merely because it describes computing nodes of varying performance. Rather, it discloses exclusively the use of CPUs, which are not boosters. Furthermore, there is no redistribution of subtasks as claimed, but merely a load balancing of homogeneous data sets. No calculations or computational tasks are disclosed, nor is any information relating to the processing of such tasks.

73 The teaching of the patent at issue is also based on an inventive step. The person skilled in the art would have had no reason to integrate boosters into the system disclosed by Li and to distribute subtasks between the nodes. With regard to Phull, even in combination with Li, not all features of the patent claim are disclosed.

74 The additional prior art submitted for the first time with the rejoinder should not be admitted. It was submitted late and the reasons given were not convincing. From the outset, the claimants had maintained an interpretation according to which the contested embodiment infringed the patent.

75 In the claimants' view, the Wan citation should also not be admitted into the proceedings. With the 'front-loaded' system of the proceedings, the UPC Agreement and the Rules of Procedure deliberately accept that prior art discovered only after all time limits have expired can no longer be taken into account. The defendants have not put forward any arguments that could justify a deviation from this general rule. Furthermore, the defendants would be unduly disadvantaged by such admission. In the short time remaining until the oral hearing, it is not possible to mount a proper defence, which might also include alternative claims.

76 Nor does Wan's teaching anticipate the subject-matter of the patent at issue in a manner that would render it non-novel. A 'subtask' according to Wan is processed in full in a single iteration. In this respect, there is no iterative processing of a task. Furthermore, there is no initial or subsequent distribution of subtasks. The only thing that happens is that the amount of data to be processed is adjusted after each iteration depending on a newly calculated task distribution ratio. Similarly, a dynamic allocation of computing nodes and booster nodes is not disclosed.

REASONS FOR THE DECISION

77 The infringement claim is admissible but unfounded. Consequently, no decision is required on the counterclaim.

A Admissibility of the infringement action

78 The infringement action is admissible. In particular, the claimants are entitled to bring proceedings before the Unified Patent Court pursuant to Article 47 of the UPC Agreement.

I. Second Claimant

79 Claimant 2) is entitled to bring proceedings before the Unified Patent Court pursuant to Article 47(1) of the UPC Agreement, as she is the proprietor of the patent at issue.

80 It cannot be successfully argued that the second claimant granted exclusive licences in respect of the patent at issue, with the result that it is no longer entitled to use the patent at issue and is therefore not affected in its own rights (see

Düsseldorf local division, decision of 13 May 2025, UPC_CFI_505/2024 – Sanofi v Amgen).

81 Article 47(1) of the UPC Agreement links the claimant's status as a party solely to the capacity of patent ownership. This is based on the notion that the patent proprietor should be entitled to assert all rights arising from their patent through legal proceedings before the Unified Patent Court. Consequently, the patent proprietor is also entitled, pursuant to Article 47(4) of the UPC Agreement, to join proceedings brought by an exclusive licensee as a party. It is not a prerequisite that the patent proprietor has personally suffered claims arising from a patent infringement, provided that only their own patent is affected. The same can be inferred from Article 47(2) of the UPC Agreement, according to which the exclusive licensee has a right of action alongside the patent proprietor only if this is not expressly excluded by contract. In light of all this, the patent proprietor is always entitled to bring an action.

II. Claimant 1)

82 However, Claimant 1) is also entitled to bring an action. Her entitlement arises from Article 47(2) of the UPC Agreement.

1. Exclusive licence

83 The first claimant holds a comprehensive exclusive licence to the patent at issue.

a)

84 On 29 August 2021, the claimants entered into an agreement whereby the second claimant granted the first claimant an exclusive licence to the patent at issue (Exhibit K 50).

85 Pursuant to Clause 11.2 of this licence agreement, German substantive law applies to the agreement.

86 The agreement was validly concluded. It was signed on behalf of both parties by their respective Chairman of the Board, Bernhard Frohwitter. According to the submitted extracts from the commercial register, he was and is authorised to act as sole representative for both Claimants (Exhibits K 1, K 2 and K 66). Furthermore, contrary to Section 181 of the German Civil Code (BGB), he was permitted to sign the agreement on behalf of both contracting parties simultaneously. The defendants were right to drop the objection of a self-dealing transaction. This is because the extracts from the commercial register submitted show that the Chairman of the Board, Frohwitter, is exempt from the prohibition on self-contracting on behalf of both Claimants.

87 The licence agreement is not void under Section 117(1) of the German Civil Code (BGB). Even though the licence was granted free of charge, it was not granted for appearance's sake. There is neither a legal principle nor even a presumption that a declaration of intent, by which one party is promised a

performance without synallagmatic consideration, was made merely for appearance's sake. Nor do the defendants put forward any further evidence to support such an assessment. The fact that the declarations were not made for appearance's sake, but that the legal consequence of a valid grant of licence was in fact intended, is also evident from the fact that the first claimant is acting as a claimant in the present case. She can only do so if an exclusive licence has been validly granted to her.

The content of the licence agreement dated 29 August 2021 is not in dispute between the parties. According to this, the exclusive licence under Clause 4.1 of this agreement was intended to cover, in substance, the entire scope of the inventions protected by the contractual intellectual property rights – including the application for the patent at issue – in particular microelectronics. Only the scope of application of the system architecture of supercomputers (high-performance computing), including cloud computing and the integration of quantum computers into HPC environments, was to be excluded from the licence.

b)

88 The second claimant further granted the first claimant, via FL Systems AG & Co. KG, an exclusive licence to the patent in suit for the remaining scope of application, so that the first claimant is comprehensively and exclusively licensed.

aa)

89 On 21 December 2023, the second claimant entered into a licence agreement with FL Systems AG & Co. KG, whereby the latter was granted an exclusive licence to the patent in suit for the scope of application relating to the system architecture of supercomputers (High Performance Computing), including cloud computing and the integration of quantum computers into HPC environments (Exhibit K 48).

90 The agreement was validly concluded. Pursuant to Clause 12.2 of the agreement, this is governed by German substantive law. The Chairman of the Management Board of Claimant 2), Bernhard Frohwitter, signed the agreement on behalf of Claimant 2) as a contracting party and, at the same time, on its behalf as general partner of FL Systems AG & Co. KG. As evidenced by the extracts from the commercial register of both contracting parties, Mr Bernhard Frohwitter was, at that time, authorised to act as sole representative for the second claimant and FL Systems AG & Co. KG, and was also exempt from the prohibition on self-dealing under Section 181 of the German Civil Code (BGB) (Annexes K 1, 63 and 64). Nor do the defendants contest this.

91 As regards the defendants' objection that this licence agreement is void pursuant to Section 117(1) of the German Civil Code (BGB) on the grounds that it constitutes a sham transaction due to the absence of consideration, this cannot be accepted. For the reasons, reference is made without restriction to the arguments concerning the licence agreement between the Claimants dated 29 August 2021.

bb)
For its part, FL Systems AG & Co. KG granted Claimant 1 an exclusive licence within the scope of the agreement of 21 December 2023 by agreement dated 7 June 2024 (Exhibit K 49). The comments made regarding the two preceding agreements apply to this agreement. German substantive law applies to it. The contract came into effect upon signature by the Chairman of the Management Board of Claimant 1) and Claimant 2) as general partner of FL Systems AG & Co. KG, as Mr Bernhard Frohwitter was authorised to act as sole representative and was exempt from the prohibition on self-dealing (see Commercial Register extracts K 2, 63 and 64). The defendants were correct not to raise the defence of nullity on the grounds of a sham transaction. The contracting parties even agreed on consideration.

2. Notification of the patent proprietor

It is undisputed between the parties that the second claimant, as the patent holder, was notified of the infringement action brought by the first claimant within the meaning of Article 47(2) of the UPC Agreement, insofar as this is necessary at all when both parties bring an action jointly from the outset.

B Merits of the infringement action

92 The infringement action is unfounded.

I. State of the art, problem and solution

93 The invention according to the patent at issue relates to a mechanism for performing computational tasks within a computing environment, in particular a heterogeneous computing environment, which is designed for the parallel processing of a computational task.

94 The patent at issue states, with reference to the prior art, that the invention claimed in the patent at issue constitutes a further development of the system described in the earlier application WO 2012/049247 A1 (hereinafter: WO'247, submitted in these proceedings as Annex BP 4), which describes a cluster computer architecture comprising a plurality of computing nodes and a plurality of boosters connected to one another via a communication interface. A resource manager is responsible for dynamically allocating one or more of the boosters and the compute nodes to one another during runtime. An example of such dynamic process management is disclosed in Clauss et al. ‘Dynamic Process Management with Allocation-internal Co-Scheduling towards interactive supercomputing’, COSH 19 January 2016, Prague, CZ. According to the patent at issue, whilst the order disclosed in WO ’247 does indeed provide a flexible order for the allocation of boosters to compute nodes, it does not address the question of how tasks are handled

are to be distributed between compute nodes and boosters (para. [0002]; paragraph references without a source citation are those of the patent at issue, Annex K 44).

95 The order described in WO '247 is also described by Eicker et al. in “The DEEP Project: An alternative approach to heterogeneous cluster computing in the many-core era”, Concurrency Computing: Practice and Experiment 2016; 28:2394–2411. According to this, the heterogeneous system comprises a large number of compute nodes and a large number of booster nodes, which are connected via a switchable network. To process an application, the application is ‘taskified’ to indicate which tasks can be offloaded from a compute node to a booster. This division of tasks is achieved by the application developer annotating the code with ‘pragmas’ that indicate dependencies between different tasks, as well as with ‘labels’ that mark the highly scalable code segments to be processed by a booster. In this context, scalability means that as the service’s workload increases, the same level of service per user can be maintained through an incremental and linear increase in hardware (para. [0003]).

96 The patent at issue also cites US 2017/0262319 A1, which, in one aspect, describes runtime processes that could fully or partially automate the distribution of data and the allocation of tasks to computing resources. A so-called tuning expert, i.e. a human operator, may still be required to assign actions to the available resources. Essentially, it describes how a specific application to be computed can be mapped onto a specific computing hierarchy (para. [0004]).

97 No specific task is formulated in the patent at issue. However, against the background of the prior art described, the technical problem can be described as providing an improved method for operating a heterogeneous computing system and such a heterogeneous computing system in order to compute a computational task more efficiently (para. [0017]).

98 To solve this problem, the patent at issue proposes a method having the features of patent claim 1, which can be subdivided as follows:

1. A method for operating a heterogeneous computing system (10) comprising a plurality of computing nodes (20) and a plurality of booster nodes (22);
2. at least one of the plurality of compute nodes (20) and the plurality of booster nodes (22) is configured to compute a computational task, wherein the computational task comprises a plurality of subtasks;

3. in a first iteration of the computation, the plurality of subtasks is allocated in a first distribution to one of the plurality of compute nodes (20) and booster nodes (22) and processed by the same;

4. information

4.1 relating to the processing of the plurality of subtasks by the plurality of computing nodes (20) and booster nodes (22),

4.2 are used to generate a further distribution of the subtasks among the computing nodes (20) and booster nodes (22) for processing in a further computational iteration.

99 The computing system disclosed further, having the features according to independent claim 9 of the patent at issue, is not relevant to the dispute.

II. Interpretation

100 Claim 1 of the patent at issue relates to a method for operating a heterogeneous computing system. The computing system comprises a plurality of computing nodes and booster nodes for computing a computing task comprising a plurality of subtasks. The method comprises, in a first step, the allocation of the plurality of subtasks to the computing nodes and booster nodes and their processing by these, and, in a second step, the use of information relating to the processing of the subtasks by the nodes to generate a further distribution of the subtasks among the nodes. The details of claim 1 require interpretation insofar as they are relevant to the decision in the legal dispute.

1. Principles of interpretation

101 Under Article 69 EPC and the Protocol on its interpretation, a patent claim is not merely the starting point but the decisive basis for determining the scope of protection of a European patent. The interpretation of a patent claim does not depend solely on the strict, literal meaning of the wording used. Rather, the description and the drawings must always be consulted as aids to the interpretation of the patent claim, and not merely to clarify any ambiguities in the patent claim. However, this does not mean that the patent claim serves merely as a guide and that its subject-matter also extends to what, after examination of the description and the drawings, appears to be the subject-matter for which the patent proprietor seeks protection (Court of Appeal, Order of 23 February 2023, UPC_CoA_335/2023; Order of 13 May 2024, UPC_CoA_1/2024; Central Chamber, Munich, decision of 16 July 2024, UPC_CFI_1/2023; Local Division, Paris, decision of 4 July 2024, UPC_CFI_230/2023; Local Division, Munich, decision of 31 July 2024,

UPC_CFI_233/2023; Hamburg local division, 26 August 2024, UPC_CFI_54/2023; Düsseldorf local division, decision of 10 October 2024, UPC_CFI_363/2023; decision of 31 October 2024, UPC_CFI_373/2024).

102 A patent claim must be interpreted from the perspective of a person skilled in the art. When interpreting a patent claim, the person skilled in the art does not apply a philological understanding, but determines the technical meaning of the terms used with the aid of the description and the drawings. A feature in a patent claim must always be interpreted in the light of the claim as a whole (Court of Appeal, UPC_CoA_1/2024, decision of 13 May 2024). The technical function that these features actually have, both individually and collectively, must be derived from the function of the individual features in the context of the entire patent claim. It may be apparent from the description and the drawings that the patent specification defines terms independently and, in this respect, constitutes a lexicon specific to the patent. Even if the terms used in the patent deviate from general linguistic usage, it may therefore be the case that ultimately the meaning of the terms as derived from the patent specification is decisive (Central Chamber, Munich, decision of 16 July 2024, UPC_CFI_1/2023; Düsseldorf local division, decision of 31 October 2024, UPC_CFI_373/2024).

2. Person skilled in the art

103 The parties agree that the average person skilled in the art referred to in the patent in suit holds a Master of Science degree or a degree in engineering in the field of electrical engineering or computer science. Furthermore, the Court assumes that the person skilled in the art has several years' professional experience in the field of high-performance computing. Insofar as the defendants suggest that the person skilled in the art should have professional experience specifically in the design and development of solutions for the calculation and distribution of computational tasks in a heterogeneous computing environment, this appears to be too narrowly defined. However, this specific technical experience is encompassed as a sub-aspect of professional experience in the field of high-performance computing.

3. Computational task and subtasks

104 According to feature 1, the method of claim 1 of the patent at issue is intended for the operation of a heterogeneous computing system comprising a plurality of computing nodes and a plurality of booster nodes. According to feature 2 of the disputed patent claim 1, at least one of the plurality of computing nodes and the plurality of booster nodes is to be configured to compute a computational task comprising a plurality of subtasks. It is therefore only required that one of the nodes is capable of computing a computational task comprising a plurality of subtasks.

a)

105 Feature 2 makes no statement as to whether booster nodes are allocated to the computing nodes for the calculation of the computational task. Nor is there any indication of a dynamic allocation of booster nodes to computing nodes, as was known in the prior art from WO'247 (see para. [0002]). This is not reflected in claim 1 of the patent at issue. Features 3 and 4, too, relate only to the allocation of the subtasks to the plurality of computing nodes and booster nodes, but not to the allocation of nodes to one another.

106 In this regard, it must be borne in mind that claim 1 of the patent at issue is a method claim. Feature 2 is formulated in such a way that it sets out requirements for the heterogeneous computing system comprising computing and booster nodes, which is to be operated by the method according to the patent. It must be capable of computing a computational task comprising a plurality of subtasks. Feature 2 constitutes a method step, if at all, only insofar as the computational task is actually computed with regard to features 3 and 4. Otherwise, however, neither feature 2 nor the remainder of the claim contains a method step in which computational nodes are allocated to booster nodes or vice versa.

107 However, the patent at issue does not preclude such a dynamic allocation of computing and booster nodes. Sub-claim 4 relates to such an allocation, which is also described in the example of the patent at issue (lines 27–29 of paragraph [0018]). Claim 1 of the patent at issue does not, however, mandatorily provide for such a type of allocation.

b)

108 The patent at issue does not contain an exhaustive definition of the term 'computational task' and its 'sub-tasks'. A person skilled in the art will understand a computational task within the meaning of the patent at issue to be one that contains a working or computational instruction suitable for a processor, in the sense of program code, which can be divided into subtasks. In this respect, it differs from a mere data set that is input into an application present on a processor (or otherwise loadable).

109 Instead, the computational task is structurally comparable to the application processing the data, even if it is typically more complex when it is to be processed on a heterogeneous cluster computer system. In any event, the patent at issue itself also uses the term 'application' in the context of describing the prior art, from whose code tasks can be derived by tagging the code ('taskified') to indicate relationships between individual tasks and to highlight scalable tasks (para. [0003]).

110 As an example of a computational task, the patent at issue cites a ‘Monte Carlo’ simulation in which an effect is modelled using a random number, with the calculations being repeated many times in succession (para. [0015]). The computational task here therefore consists of an algorithm that is applied countless times using different random numbers.

c)

111 Insofar as the parties dispute whether a computational task must necessarily be heterogeneous, i.e. whether it must be divisible into subtasks that are suitably suited for solution by a compute node or a booster node, or whether it may consist solely of homogeneous (uniform) subtasks, it must be borne in mind that the subject-matter of the claim is not the computational task itself, but a method for operating a heterogeneous computing system. The method and the computing system do not in themselves determine the computational task, provided that it is in fact a task and not merely a data set or the like. It is therefore not excluded that, using the protected method on a heterogeneous computing system, a computational task may be computed which comprises entirely uniform sub-tasks that are particularly suited only for processing by one of the node types. However, it follows from the very concept of a heterogeneous computing system that, due to the qualitative technical differences between computing nodes and booster nodes, the method must also be capable of computing heterogeneous computational tasks

4. Iteration

112 In accordance with features 3 and 4, the multitude of subtasks are the subject of a first and a further iteration or computing iteration. This refers to the repeated execution of the computational instruction constituting the subtask or the computational task.

a)

113 Neither claim 1 nor the description of the patent at issue defines the term ‘iteration’. This is assumed by the patent at issue. According to the general understanding of the term, the term ‘iteration’ describes the repetition of a process, usually in order to improve it or its result (see p. 23 of the defendant’s rejoinder of 13 August 2025, with reference to the Cambridge Dictionary). Specifically, in a mathematical sense, an iteration describes the repeated application of the same computational instruction.

b)

114 The description of the patent at issue also refers to precisely this meaning. It must be borne in mind that the patent-protected method is fundamentally directed towards the calculation of complex computational tasks on heterogeneous cluster computers. This includes, in particular, simulations such as those covered by the patent at issue with the ‘Monte Carlo’ simulation

is cited by way of example (see para. [0015]). It is expressly stated there that the calculations are repeated many times in succession (“the calculations being repeated many times in succession”, lines 51 ff. in para. [0015]). The aim is therefore to execute the computational instructions with ever-changing initial values or on the basis of modified data sets, thereby representing, for example, an approximate value, a limit value or a trend over time. In this context, the results of individual subtasks may also serve as the initial values for the calculation of other subtasks or even the same subtask (loop).

115 In so far as the claimants argued during the oral hearing that the cited passage (para. [0015]) does not in fact use the term ‘iteration’ and that an iteration within the meaning of the patent at issue does not in fact require the repetition of the calculation, this argument cannot be accepted. For the interpretation of the term ‘iteration’, it is not necessary for it to be specifically mentioned or even defined in the cited passage. It is sufficient if the person skilled in the art can deduce the technical meaning of the term used in the patent claim with the aid of the description and the drawings. That is the case here. The cited passage is the only part of the description of the patent at issue that provides a concrete example of a computational task and its calculation. The description is consistent with the general technical understanding of the term ‘iteration’ held by a person skilled in the art, and there is no apparent reason to deviate from this understanding.

c)

116 Even from the required functional perspective, this understanding of the term ‘iteration’ technically sensible. In this respect, it must be borne in mind that claim 1 is a method claim relating to the computational task. The computational task, with its sub-tasks, and the manner of its calculation are therefore not part of the technical solution to the problem underlying the patent at issue, but rather its starting point.

117 The technical objective of the patent at issue is to improve the efficiency of computing a computational task using a heterogeneous computing system (para. [0017]). According to features 3 and 4, this is achieved by a specific redistribution of the sub-tasks among the computing nodes and booster nodes on the basis of information regarding the processing of the sub-tasks in (at least) one previous iteration (see also paragraphs [0006], [0017] and [0018]). The information indicates whether other computing or booster nodes are more suitable for processing the sub-tasks. In this respect, the method according to the patent presupposes the iterative computation of the sub-tasks. For it is precisely because the sub-tasks are computed repeatedly that the information obtained from past iterations in accordance with feature 4 allows a statement to be made about future processing by a computing or booster node. The information according to feature 4 therefore relates to the processing of the subtasks by the plurality of computing and booster nodes and not to the subtask or the

node as such (see below for further details). The processing of the subtasks by a specific node provides the information from which it can be deduced whether the calculation of the subtask, repeated in a further iteration, is more efficient when performed by the same or a different node. However, the patent at issue does not provide a solution as to how the computation of the computational task could be optimised if the computational instructions defined by the subtasks were not repeatedly computed. It does not explain to what extent information about the past processing of computational instructions could contribute to the future processing of other (since non-repeated) computational instructions.

118 Consequently, the claimants' view that, given the required functional interpretation, there is no need to wait until the completion of an initial 'iteration' before proceeding with a further allocation of subtasks—particularly where individual subtasks are highly complex and take considerably longer to process than simpler ones—cannot be accepted. In such a case, it cannot be ruled out that the calculation may be interrupted and the initial allocation reviewed. An iteration is therefore the processing step or the processing time after which an assessment of the system's status and the initial allocation can reasonably be carried out, and meaningful information is available for further, essentially system-wide processing or allocation.

119 Neither the concept of iteration nor the patent at issue provides any basis for such an interpretation. The claimants reduce the concept of iteration to the stage of processing at which an evaluation of the task and a reallocation of subtasks would appear appropriate. However, this has nothing to do with the actual meaning of the term 'iteration' in the sense of a repetition. Nor is there any indication that, according to the teaching of the patent at issue, the iteration is intended to have the function or even the significance claimed by the Claimants. Nothing in the patent at issue suggests that an iteration should in any way depend on when an examination of the initial distribution and a further distribution appear appropriate. On the contrary, the iteration arises as such from the nature of the computational task in the sense of an iterative computational instruction, and it is only the repeated application of the computational instruction that enables – as explained – a more efficient distribution across the compute and booster nodes for a further iteration – understood here as the repeated execution of the same computational instruction.

d)

120 It must also be taken into account that the second claimant, in its response to the EPO's communication in the grant proceedings for the patent at issue, took a different view and proceeded on the basis of a correct understanding of the concept of iteration (Annex BP 7). Such statements by a patent applicant in the grant proceedings do not, however, constitute interpretative material for the interpretation of the granted patent within the meaning of Art. 69

(1) EPC. Nor are such statements by the applicant in the grant proceedings binding for the purposes of interpretation in subsequent infringement or nullity proceedings. However, they may provide an indication of the skilled person's understanding of the teaching of the granted patent at the priority date (Court of Appeal, Order of 20 December 2024, UPC_CoA_402/2024 – Alexion v Samsung Bioepis), because the applicant themselves usually has the best understanding of their invention and the relevant technical knowledge.

121 In its response to the prior art cited against it, the second claimant stated that a computational task could, for example, be divided into a first part and a second part, and that the second part could be computed in parallel with the first part or after the first part. However, this does not disclose a multitude of subtasks that are computed in several iterations (Annex BP 7, second paragraph on p. 2). According to the view originally expressed by the second claimant, the iterative calculation of the computational task or its subtasks thus goes beyond the mere division of a computational task into subtasks and the parallel or successive calculation of these subtasks. A parallel calculation is ruled out simply because the first and subsequent iterations must necessarily take place one after the other in time. However, a merely successive calculation of subtasks of a computational task is also insufficient, as it lacks the repeated execution of the same computational instruction. Instead, an iteration is aimed at executing the computational instruction contained in the subtask repeatedly – this time with modified values – in order to derive further insights from the various results of the repeated application, which consequently go beyond the result of the one-off calculation of the subtask.

5. Initial and subsequent allocation of subtasks

122 If the computational task, together with its sub-tasks to be calculated iteratively, forms the starting point of the method, these sub-tasks are now, in a first method step in accordance with feature 3, allocated to some of the plurality of computational and booster nodes and processed by them. In a second method step in accordance with feature 4, a further distribution of the sub-tasks among the computational and booster nodes is generated.

a)

123 The further distribution according to feature 4 does not necessarily have to encompass all the subtasks that were also the subject of the first distribution. According to the teaching of the patent at issue, it is permissible for only a portion of the original subtasks to be redistributed. This, too, constitutes a further distribution of the subtasks amongst the computing and booster nodes. It is also possible for several subtasks, which were processed by different nodes, to be computed by one and the same node in a further distribution.

124 The wording of features 3 and 4 does not necessitate an interpretation whereby all subtasks that were already the subject of the first distribution are also covered by the further distribution. The fact that feature 4, with the further distribution of ‘the subtasks’, refers linguistically back to ‘the plurality of subtasks’ according to feature 3 does not mean that exactly the same set of subtasks must be redistributed. Linguistically, the claim also permits an interpretation according to which only some of the subtasks are redistributed, because this too constitutes a further distribution of the subtasks (whilst leaving a part with the nodes to which they were originally allocated).

125 It should be borne in mind that, during the first computational iteration, it may become apparent that certain subtasks are not suitable for processing by specific nodes. It seems appropriate to redistribute only these subtasks. This is precisely the purpose of using the information regarding the processing of the plurality of subtasks by the plurality of nodes in feature 4. It is not clear why subtasks that are processed by a suitable node and for which there is otherwise no reason for an allocation to another node should be reassigned.

126 This is precisely what follows from the description of the patent at issue. Accordingly, certain – and thus not all – sub-tasks (‘certain sub-tasks’) which were allocated to a booster in an initial distribution may prove unsuitable for processing by the booster, such that processing the sub-tasks by a computing node instead of a booster could optimise the calculation of the task as a whole (para. [0017]). The patent at issue accordingly proposes a further iteration of the task with a modified distribution of sub-tasks in order to improve the efficiency of the task’s computation (para. [0017]).

127 The description of the patent at issue notes that the (further) distribution of subtasks may also be influenced by the interdependence of subtasks. If, for example, a first sub-task being processed by a booster requires inputs from a second sub-task that is not being processed by the booster, this may – according to the patent specification – lead to an interruption in the processing of the first sub-task. Accordingly, following a further distribution in a subsequent iteration, both the first and the second sub-task may be processed by the booster (para. [0022]). Here, too, only the reallocation of one of the sub-tasks is required with the further distribution.

128 In light of the above, neither the greater suitability of a computing or booster node for processing a subtask, nor the associated increase in efficiency when calculating the computational task, necessitates the reallocation of all subtasks that were the subject of the first distribution in the subsequent distribution.

b)

129 However, the decisive factor is that the patent at issue provides for the further distribution of subtasks between the computing and booster nodes after the first iteration in order to solve the technical problem. The further distribution according to feature 4 enables the allocation of sub-tasks to compute or booster nodes that are better suited to processing these sub-tasks, so that the sub-tasks and thus the entire computational task can be processed more efficiently. This solves the technical problem underlying the patent at issue, namely to improve the operation of a heterogeneous computing system, in particular to optimise the processing of computational tasks (see para. [0017]).

130 The teaching of claim 1 of the patent at issue does not preclude individual subtasks from being solved and completed within one of the computational iterations, or even from being redefined in terms of their content. However, this is not the subject-matter of the invention, nor is it the solution to the technical problem. Instead, the invention is directed towards redistributing the set of sub-tasks to be calculated iteratively for a specific computational task – a set which is present from the outset and remains essentially unchanged – thereby enabling improved processing of the sub-tasks. It cannot be inferred from the patent at issue that the computational task is divided into new sub-tasks, which are then the subject of the further distribution.

6. Information relating to the processing of the plurality of subtasks 131 According to the teaching of the patent at issue, the further distribution according to feature 4 is to be generated on the

the use of information relating to the processing of the plurality of subtasks by the plurality of computing nodes and booster nodes. The information must therefore be causal for the further distribution.

a)

132 As is already apparent from the wording of feature 4, the information referred to therein must be such that it provides information about the sub-tasks themselves and their actual processing by the computation and booster nodes, whereby such processing is contingent upon the first or, in any event, preceding computation iterations. It is therefore not sufficient for the information to relate to the computing and booster nodes as such and, for example, to reflect their activity, utilisation, capacity or the like (hereinafter referred to as ‘status’ or ‘status of the nodes’).

b)

133 This interpretation also follows from the description of the patent at issue, which expressly distinguishes between information relating to the processing of the subtasks and the status of the nodes. For instance, regarding an embodiment, it states:

> “Accordingly, the system 10 includes a mechanism whereby each of the computation nodes and the cluster nodes are arranged such that daemons 26a, 26b and 32 provide feedback to daemon 34 regarding the processing of sub-tasks and the current state of the respective processing entity.” (para. [0018]).

In a preferred embodiment of this example, in which a scaling factor for each sub-task was estimated by an operator and entered into the system, it is also stated:

> “During a first iteration of the task, the results of the execution of the sub-tasks are collected together with the information from the daemons concerning the processing of the sub-tasks and the status of the nodes.” (para. [0019])

134 The patent at issue thus expressly distinguishes between information relating to the processing of the sub-tasks and the status of the nodes. Both are to be communicated by the nodes’ daemons to the application manager in order to decide, on this basis, on a reallocation of the sub-tasks. However, only the information relating to the processing of the sub-tasks has been received in claim 1 of the patent at issue. This does not preclude the status of the nodes from also being communicated and taken into account in the further distribution of the subtasks. What is decisive, however, is that a distinction must be made between the information concerning the processing of the subtasks and the status of the nodes, and that the former must, according to the teaching of the patent at issue, be used for the generation of the further distribution.

135 Nor can a different interpretation be justified by the fact that the description goes on to state:

> “For each iteration, the daemons operating in the computation nodes and the boosters report status information to the application manager and the resource manager, enabling the calculation of subsequent iterations to be optimised by further adjustments to the allocation of sub-tasks to computation nodes and boosters.” (para. [0020])

Insofar as this passage refers to status information that is reported, it merely takes up the previously described embodiment. The status information is to be understood as the information concerning the processing of the sub-tasks, i.e. the processing status, and the status of the nodes. There is no indication that the patent at issue refers here exclusively to the status of the nodes to be reported, or that the information mentioned in feature 4 of claim 1 can comprise solely this status of the nodes. This is already contradicted by the wording of the feature, which relates solely to information concerning the processing of the plurality

of sub-tasks (‘information relating to the processing of the plurality of sub-tasks’).

136 The information relating to the processing of the sub-tasks may include information on the suitability of the sub-tasks for processing by the corresponding computing or booster node (paragraphs [0017] to [0019]), such as the scalability of the sub-task (see lines 5–7 of paragraph [0022]). However, information arising in the course of processing a sub-task, for example because the calculation of one sub-task depends on another sub-task, is also to be regarded as information within the meaning of feature 4 which relates to the processing of the sub-tasks and which may determine further distribution. In the specification of the contested patent, this is described as follows for an embodiment:

> “(…) the distribution may also be influenced by information learned about the processing of the sub-task and any need to call further sub-tasks during processing. If a first sub-task being handled by a booster requires input from a second sub-task not being handled by the booster, this may lead to an interruption in the processing of the first sub-task. Accordingly, the daemon at the booster handling the first sub-task can report this situation to the application manager such that in a further iteration both the first and second sub-tasks are handled by the booster.”

137 The cited embodiments also make it clear that the information within the meaning of feature 4 does not merely concern the general status of the processing of all sub-tasks, but originates from the processing of the individual sub-task by the computing or booster node to which it is allocated, and that further distribution is generated on the basis of this information. In this respect, it is not necessary for all subtasks to be redistributed, but only those for which the need arises from the respective information.

c)

138 It cannot be successfully argued against this interpretation that the information referred to in feature 4 of claim 1 of the patent at issue also includes status information of the nodes, because the patent at issue recognised, in contrast to the prior art, that the processing of the subtasks has a further-reaching influence on the status of the nodes and does not merely concern their utilisation, if, for example, the suitability of the node for the subtask is also taken into account. This line of argument alone demonstrates that the information must necessarily be based on the processing of the subtask as such, so that the teaching of the patent at issue can at all be distinguished from the – undisputed – load balancing between the nodes based on their status information, which is known in the prior art. If the inventive concept lies precisely in utilising the information concerning the processing of the subtasks, this must also be the subject matter of the claim. This is precisely what follows from the wording of feature 4.

139 This reasoning also serves to counter the claimants' conclusions drawn from the argument that the insight underlying the patent at issue consists in placing the consideration of load distribution, previously known in the prior art, in the context of the actual processing of the subtasks for the first time. Even if this were the case, it cannot be concluded that the load distribution or the other status of a node is sufficient as information relating to the processing of the sub-task. For this is not information that specifically concerns the processing of the respective assigned sub-task.

140 The statement made by the second claimant during the grant proceedings cannot be interpreted otherwise; it constitutes a further indication of the interpretation. In explaining her understanding of the information within the meaning of feature 4, the second claimant refers precisely to the previously cited embodiments of the patent at issue and, in distinction from the prior art, limits the information to that relating to the scalability of the sub-task and that obtained from the processing of the sub-task and the necessity of calling up further sub-tasks during processing. The second claimant explicitly states that the invention according to the patent at issue relates to the analysis of the subtasks and the obtaining of information relating to the processing of the plurality of subtasks, and not to mere load balancing (Annex BP 7, third paragraph on p. 2).

d)

141 In so far as the claimants argued for the first time at the oral hearing that features or characteristics of a computational task which had been defined in advance by the user were also to be regarded as information within the meaning of feature 4, this cannot be accepted.

142 According to the wording of the claim, the information under feature 4 must relate to the processing of the plurality of subtasks by the plurality of computing and booster nodes, and not merely to the computational task or the subtasks as such. The teaching of the patent at issue consists precisely in deriving information from the processing of the sub-tasks in preceding iterations, which makes it possible to recognise that the sub-tasks can be executed more efficiently on other computing and booster nodes, and to redistribute the sub-tasks accordingly. Features or characteristics of the computational task or its subtasks specified in advance by the user are inconsistent with this, as they make no statement regarding the actual processing by the computational or booster node to which the subtask is allocated in the first iteration. There is no information regarding the processing of the subtask that could be causal for the further distribution of the subtask.

143 Nor does anything else follow from the description of the patent at issue, according to which a user, prior to task processing, defines a scalability factor for each sub-task

estimated, which is entered into the computing system and forms the basis for the initial distribution of the sub-task (paras. [0009] and [0019]; see also para. [0003] on the prior art). This information provided by the user only influences the initial distribution and the first iteration (see ‘initial distribution’ in para. [0009]). The subsequent distribution, which is the sole focus of feature group 4, is based exclusively on information obtained from the processing of the sub-tasks in the preceding iterations (see “During a first iteration of the task, the results (…) are collected together with the information from the daemons concerning the processing of the sub-tasks (…). The application manager then performs a reassignment of the sub-tasks (…)” in para. [0019]). The information initially provided by the user therefore plays no role in the subsequent distribution.

III. Infringement

144 Regardless of whether it is offered and supplied with or without pre-installed Run:ai software, the contested embodiment is not suitable for implementing the method protected by claim 1 of the patent at issue.

1. Contested embodiment

145 The infringement action challenges the DGX System products advertised and placed on the market by the defendants, regardless of whether they are sold with or without pre-installed Run:ai software.

146 Nvidia DGX is a range of Nvidia servers and workstations which, in addition to using CPUs, are specifically designed to utilise graphics processing units (GPUs) to accelerate deep learning applications. According to the (non-exhaustive) list provided by the claimants, these DGX systems include systems and clusters such as the DGX H100, DGX A100, DGX H200, DGX GH200, DGX B200 and DGX GB200 SuperPOD. All these systems are examples of the contested embodiment because, according to the defendants’ undisputed submission regarding the allegation of infringement, they are identical.

147 The claimants stated in the statement of claim that the contested embodiment consists of DGX system products offered or distributed by the defendants, which are supplied with Run:ai software. However, after the defendants contested in their defence that the Run:ai software was not yet installed at the time of purchase of a DGX product, but had to be downloaded and installed by the customer independently, the claimants stated in their Reply that there could be no doubt of an indirect infringement, even even if the Run:ai software is not offered and supplied as part of a bundle with the DGX systems. Contrary to the defendants’ view, this submission by the claimants does not constitute an amendment or extension of the claim within the meaning of Rule

263 of the RoP, which requires a specific application and separate authorisation.

a)

148 According to the case law of the Court of Appeal, not every new argument constitutes an amendment to the claim that obliges a party to apply for leave under Rule 263 of the Rules of Procedure. An amendment to the claim occurs when the nature or scope of the dispute changes. In infringement proceedings, this is the case, for example, where the claimant asserts a different patent or objects to a different product (Court of Appeal, Order of 21 November 2024, UPC_CoA_456/2024 – OrthoApnea).

149 The nature and scope of a legal dispute are determined, on the one hand, by the applications made, but also by the facts put forward in support of those applications and the underlying claims. In line with the Court of Appeal, it must be assumed that not every additional or amended statement of facts automatically leads to an amendment of the claim. This also applies where, following the defendant's statement of defence, the product in question appears in detail to differ from how the claimant described it in the statement of claim, provided that the allegation of infringement continues to relate in fact to the product described in general terms. The claimant is not always aware of all the technical details of a product in question, so it must be possible for them to respond to clarifications and corrections provided by the defendant in the defence and to adapt their submission accordingly.

b)

150 This is also the case here. Insofar as the claimants now accept that the Run:ai software may, in some instances, be installed by the customers themselves, and base their claim on this scenario, they are not objecting to an entirely different product. They continue to challenge the DGX System products and, in the context of the indirect infringement, continue to focus on the use of the Run:ai software, with the sole difference being that it is not already installed upon delivery but is installed by the customers. The nature and scope of the legal dispute have also remained essentially identical following the Claimants' Reply.

151 The claimants' claim is based on the allegation of indirect infringement by the defendants. In their Reply, the claimants continue to pursue identical claims and demands. Insofar as the action was also to be based on direct patent infringement, this was not admitted.

152 Nor have the facts underlying the allegation of indirect infringement changed to such an extent that this could justify an amendment to the claim pursuant to Rule 263 of the RoP. The statement of facts regarding the acts of infringement within the meaning of Article 26(1) of the UPC Agreement – the offering and supply of DGX systems by the defendants – is

remains unchanged, as it is independent of whether the software is pre-installed or not. Furthermore, the suitability of the DGX systems for the use of the method by customers, as required under Section 26(1) of the UPC Agreement, is independent of whether the software is pre-installed or is only installed by the customers. This is because the claimants' claim is that the protected method can be applied using the DGX systems, including the Run:ai software installed on them. However, the suitability of the DGX systems for software installation is undisputed, regardless of where and when it takes place. In this respect, the question also arises as to whether DGX systems without pre-installed software constitute a means relating to an essential element of the invention in a manner entirely different from when the software is pre-installed. This is not discussed by any of the parties. In light of the above, the timing of the installation may, at most, influence whether the defendants knew or ought to have known, within the meaning of Article 26(1) of the UPC Agreement, that the DGX systems – now without pre-installed software – are suitable and intended for implementing the patented method. This question may arise somewhat differently if the DGX system products are not offered and delivered with pre-installed Run:ai software, but the software must first be installed by the customer. However, as this is only one of several requirements for indirect infringement, the assessment of which also depends on further circumstances of the individual case that have remained unchanged, the argument adapted in the Reply regarding the time of installation of the software is not sufficient to alter the nature or scope of the legal dispute.

2. Suitability for use of the protected method with pre-installed software

153 The contested embodiment is not suitable for the application of the method protected by claim 1 of the patent at issue within the meaning of Article 26(1) of the UPC Agreement. The fulfilment of feature 4 cannot be established.

a) Computational task and sub-tasks

154 It can be left open whether the contested embodiment can be regarded as a heterogeneous computing system whose CPUs and GPUs constitute a plurality of computing nodes and a plurality of booster nodes within the meaning of feature 1 of patent claim 1. In any event, it is undisputed between the parties that a 'workload' within the meaning of the terminology of the Run:ai software documentation can be regarded as a computational task and 'pods' within the meaning of that terminology as subtasks within the meaning of feature 2. It is also undisputed that the Run:ai scheduler performs the allocation of pods to the various resources – CPUs and GPUs, which for the purposes of the further infringement analysis may each be regarded as computing and booster nodes within the meaning of the patent at issue – in accordance with predetermined rules.

b) Information relating to the processing of the plurality of subtasks, and further distribution

155 However, it cannot be established that the contested embodiment, with the Run:ai software installed, uses information relating to the processing of the plurality of subtasks by the plurality of compute nodes and booster nodes to redistribute subtasks in accordance with feature 4. [redacted].

aa)

156 The claimants appeal to the Run:ai Scheduler and its ability to constantly recalculate resource allocation for each project and perform new allocation during the ongoing process.

157 The Run:ai software requires that the organisation or business unit be divided into departments and projects, and that individual users be allocated to these departments and projects. Workloads are submitted to the computing system at the project level. Furthermore, the software requires that the available resources be grouped into so-called node pools. Node pools are a logical grouping of two or more 'nodes' – in Run:ai terminology, these are servers – based on specific characteristics, such as hardware type. For example, a node pool may consist exclusively of CPU servers or comprise all high-performance GPU servers in a cluster. Finally, fixed quotas must be allocated to specific node pools for the projects ('reserved quota'). For example, the allocation of a GPU quota to a project based on a node pool means that the workloads submitted by that project are entitled to use a certain number of GPUs as guaranteed resources, and may use them for all types of workloads within that project. However, it is not mandatory for the workloads to always require all the guaranteed resources of a project's node pool. The workloads may also require fewer resources (see Annexes K 40 and K 47).

158 In principle, a workload can only be processed if the necessary CPU and GPU resources are available (for the exception due to the 'elasticity feature', see below). The pods of a workload are then distributed across these nodes. However, the Run:ai scheduler may also perform an allocation of more resources to a workload than it requests, and even more resources than the guaranteed quota of the associated project, provided such additional resources are available. For example, if a project has a 'reserved quota' of four GPUs, the Run:ai scheduler can perform an allocation of eight GPUs to a workload of this project if they are not otherwise in use, i.e. four GPUs more than required ('over quota').

bb)

159 The Run:ai software can now dynamically reallocate the assigned resources. The claimants initially appeal to the ‘elasticity feature’ in this regard. If this is activated and the requested resources are not available for a second workload, the Run:ai scheduler can, in its next planning cycle, withdraw resources exceeding the required quota of the first workload and perform its allocation for the second workload. In the example described above, a first workload had requested four GPUs and the scheduler had made an allocation of eight GPUs to it because four additional GPUs were available. If another workload now requires a guaranteed quota of three GPUs from the same node pool and no further GPUs are available, the Run:ai scheduler can perform the allocation of three GPUs that were assigned to the first workload to the second workload. Conversely, additional resources may also be allocated to a workload in a subsequent planning cycle should such resources become available.

160 However, it cannot be established that the contested embodiment fulfils feature 4 of patent claim 1 in the course of the dynamic reallocation. The allocation of resources in the subsequent planning cycle is not generated using information relating to the processing of the plurality of subtasks – in this case, the pods of the respective workload – by the plurality of compute and booster nodes. Instead, the availability of the nodes is the sole determining factor for the allocation of pods to the nodes. This has nothing to do with the processing of the subtasks in a previous iteration. A reallocation of resources takes place as soon as a workload utilises computing power beyond the required resources and when another workload or another project requests resources that are not otherwise available. For this, the Run:ai scheduler does not rely at all on information regarding the computation of the pods. Rather, an internal schema detailing the resources allocated to the workloads and projects currently in progress would suffice, showing the scheduler the resources currently free or available for withdrawal in each planning cycle. The same applies when a project or workload is completed and resources become available. Their allocation to a running workload is in no way related to the processing of the pods for that workload.

cc)

161 The claimants further base their claim of infringement on the ‘fairshare’ and the “fairshare balancing” that the Run:ai scheduler may apply.

162 According to the Run:ai documentation, projects may claim a share of a node pool’s unused resources exceeding the guaranteed resource quota (‘reserved quota’), known as ‘over-quota’ resources. Conceptually, the ‘fairshare’ is the sum of the project’s guaranteed resources plus the share of non-guaranteed resources in that node pool (‘reserved quota’ + ‘over quota’), which is calculated by the Run:ai scheduler.

‘Fairshare balancing’ now means that the Run:ai scheduler endeavours to provide each project with the resources to which it is entitled, using two main parameters, namely the earned quota and the earned fairshare (see Annex K 40).

163 Accordingly, a redistribution of resources may also take place in the course of “fairshare balancing”. If the resources of a node pool for one project are below the “fairshare” and those of another project are above the “fairshare”, the Run:ai scheduler shifts resources between the projects. However, even in this case, feature 4 of claim 1 of the patent at issue is not fulfilled. There is no evidence that this reallocation within the framework of ‘fairshare balancing’ takes place on the basis of information regarding the processing of the pods of the respective workloads. Rather, it is solely the value of the reserved resources and the over-allocated resources, and their availability, that is decisive. This does not concern the reallocation of subtasks within a workload, but rather of resources between different projects. The processing of the individual pods of a workload is irrelevant in this regard.

dd)

164 Nor can it be established that characteristic 4 applies specifically to ‘preemptable workloads’ (see, inter alia, p. 1 et seq. of Annex K 40a), to which the claimants referred in particular during the oral hearing. The ‘preemptability’ of a workload is the prerequisite for that workload to be able to utilise an ‘over quota’. Conversely, the resources allocated in excess of the ‘deserved quota’ – i.e. the ‘over quota’ – may also be withdrawn from these workloads if workloads requiring priority processing need these resources and there are insufficient resources available for all workloads. In the claimants’ view, in order to decide whether resources can be withdrawn from a workload, the Run:ai scheduler requires both the information that the workload is a ‘preemptable workload’ and the information that the workload is using ‘over quota’ resources. In the claimants’ view, this information constitutes information within the meaning of feature 4. The court cannot agree with this view.

165 The ‘preemptability’ of a workload does not constitute information relating to the processing of subtasks by the compute and booster nodes, as required by feature 4. Whether a workload is ‘preemptable’ is determined by the user and constitutes a property of the workload that is fixed before processing of the individual pods even begins. The mere fact that the workload is currently being processed and that individual pods may be excluded from processing due to the ‘preemptable’ property if other departments, projects or workloads require resources does not make the property information within the meaning of Feature 4. The information regarding ‘preemptability’ only acquires significance due to the fact that there are also workloads that do not receive an ‘over quota’ and cannot be interrupted, and are therefore not ‘preemptable’. Consequently, neither the processing of the pods by the CPUs and/or GPUs nor the information regarding

‘preemptability’ are the cause of the interruption of the processing of the subtasks, but rather solely the resource requirements of other workloads.

166 As regards the purported information as to which pods of the workload are part of the ‘over quota’ and whose processing may consequently be interrupted, it cannot be established that feature 4 is fulfilled either. It remains unclear at what level it is actually determined which pods of a workload are processed using resources allocated as ‘over quota’ and which are processed within the ‘deserved quota’. It does not even seem out of the question that the information is limited to which specific resources have been allocated to a workload as ‘over quota’. The specific pod and its processing by these resources are then irrelevant, meaning that this does not constitute information within the meaning of feature 4. In any event, there is no indication that a pod for which it was not already clear at the time of allocation that it would be processed ‘over quota’ is selected ad hoc from among several pods on the basis of its previous processing in order to terminate its processing due to other resource requirements. As the claimants themselves explained at the oral hearing, this is at most a characteristic of the task, and thus of the pod, when asked what kind of task with what characteristics is currently taking place on this or that GPU. However, the characteristic of a subtask does not constitute information relating to its actual processing by the compute and booster nodes.

167 It should also be noted, however, that the withdrawal of resources does not result in a redistribution of pods; rather, the processing of pods that have been processed ‘over quota’ is simply terminated. Insofar as the claimants appeal to the fact that the processing of these pods is resumed at a later point in time – possibly even after several scheduling cycles – once sufficient resources are available again, this does not result in a distribution within the meaning of feature 4. The information that led to the interruption of the processing of these pods is not causal for the allocation of resources. This allocation is not conditioned by information relating to the earlier processing of the pods by the respective CPU or GPU. Rather, the cause is solely the fact that further resources are available. The concept underlying the contested embodiment is, in this respect, entirely different from that provided for by the teaching of the patent at issue.

168 According to the patent-protected technical teaching, information relating to the processing of the subtasks by the nodes – and thus necessarily to the processing in earlier iterations – is used to generate a further distribution. The Run:ai software, on the other hand, aims at the fair distribution of scarce resources among different users whilst fully utilising the available resources. For the latter, information is required, at most, as to whether a CPU/GPU is available or can be made available. That this may inevitably also depend on information as to whether a CPU/GPU is in use at all

However, the fact that a sub-task is being processed and the type of workload to which it belongs (elastic/non-elastic, preemptable/non-preemptable) does not in itself make this information relevant within the meaning of feature 4.

169 Consequently, the claimants' reference to the possibility of 'bin-packing' is also inconclusive. In this variant of the software application, the Run:ai scheduler endeavours to avoid a fragmented load on the CPUs and GPUs and to move pods between the processors in such a way that a node is utilised as fully as possible and sufficient GPUs are free to process the next workload in the queue (see p. 4 of Annex K 40a under 'Bin-packing & Consolidation'). As previously explained, it cannot be established that information relating to the processing of the pods by the respective CPUs and GPUs is required for this purpose.

ee)

170 Nor does the priority of a workload constitute information within the meaning of feature 4. Priorities are assigned by users. They result in workloads with higher priority being scheduled for processing before those with lower priority. Workloads with higher priority may even displace those with lower priority within the same project (see Annex K 40). Consequently, if resources are withdrawn from one workload and allocated to another, this has nothing to do with the processing of the pods of the respective workload, but is based solely on the independent priority information, which is linked not to the pods but to the workload.

ff)

171 Finally, feature 4 of claim 1 of the patent at issue is also not realised by inference workloads and the use of the 'Hugging Face' inference model.

172 The claimants first based their allegation of infringement on this feature of the Run:ai software in their reply of 13 June 2025. Whether this submission can be disregarded pursuant to Rule 9.2 of the RoP is ultimately irrelevant. Even if it is taken into account, the use of the subject-matter of the patent at issue has not been conclusively demonstrated.

173 It has not been argued, nor is it otherwise apparent (see Annex K 56 on the provision of inference workloads by 'Hugging Face'), that this feature relates at all to the allocation of pods to compute and booster nodes, or to what extent the allocation of resources to projects and workloads differs from the previously described functioning of the Run:ai scheduler.

174 In this regard, the claimants merely appeal to the Run:ai software documentation (Annex K 56) and argue that every time a variable condition relating to the underlying inference workload (i.e. the set of subtasks) is met, Run:ai creates a new replica pod (i.e. a new subtask), to which resources – computing and booster nodes – are then allocated. The variables in question are

latency, throughput and parallelism, and thus information within the meaning of feature 4 of the patent claim.

175 This explanation falls short. This is because the variables do not serve to generate a distribution of subtasks, but rather represent a condition for the creation of a new replica. A replica is a copy of a workload or a pod that is created when a specific threshold value—set by the user—for a selected variable is exceeded. Selecting a variable and setting a threshold value are only necessary at all if the user has optionally set a minimum and a maximum value for the number of replicas to be scaled up or down, and these values are different. Only then is a replica created each time the selected condition – the value of the variable – is met (see Annex K 56).

176 However, the creation of one or more replicas does not involve any (further) distribution of the pods of the ‘Hugging-Face inference workload’ within the meaning of feature 4 of the patent claim. Rather, merely a replica—that is, a copy of a pod—is created, and resources are allocated to this replica. Of the multitude of subtasks that were the subject of the initial distribution, no subtask is redistributed by the creation of a replica within the framework of the Hugging-Face inference model. All pods continue to be processed by the nodes to which they were originally allocated. The value of the variable, which the claimants regard as information within the meaning of feature 4, is therefore not used to generate a redistribution of the pods in accordance with feature 4. No such further distribution takes place. Instead, upon fulfilment of the condition, one or more replicas are generated to which resources must be allocated. The patent at issue does not address the creation of further subtasks. In any event, however, this is not the technical teaching by which the patent at issue seeks to solve the underlying technical problem. This teaching consists in the further distribution of the original subtasks of the computational task depending on information regarding their processing in previous iterations, but not in replicating existing subtasks.

c) Iteration

177 Furthermore, it cannot be established that the contested embodiment, with the Run:ai software installed, is capable of calculating subtasks over several iterations and, following a first iteration in accordance with feature 3, redistributing the subtasks a further time in accordance with feature 4.

178 The claimants have merely argued that every planning cycle of the Run:ai scheduler constitutes an iteration within the meaning of the patent at issue. It is undisputed that the calculation of the workload can be interrupted and that the resources allocated for the calculation can be changed during this interruption. These interruptions characterise the planning cycle of the Run:ai scheduler.

179 This argument falls short. If the interpretation is correct, the calculations between two planning cycles cannot simply be equated with an iteration within the meaning of features 3 and 4. The claimants' argument does not make clear what the pods actually concern and what is being calculated.

180 The defendants have also disputed that [redacted]. They further argue that the interruption of the calculation of the pods does not result in the pods being calculated in multiple iterations. This is because the interruption does not occur after a complete calculation/iteration of the pods.

181 It therefore appears that, following an interruption, the pods are recalculated, but the calculation instruction is not executed again. In any event, it cannot be established to the satisfaction of the court that the pods are calculated in multiple iterations within the meaning of the patent at issue and redistributed after each iteration. This remains true even when taking into account that the pods relate to tasks of different kinds and that, at the time of redistribution following a planning cycle, some pods may have been completed whilst others relate to more extensive tasks that have not yet been completed. Only the latter are eligible for further distribution in accordance with feature 4. Even if these pods relate to tasks of an iterative nature, the iteration is nevertheless unrelated to the further distribution of this pod, because the distribution – according to the defendants at the oral hearing – takes place solely on the basis of fixed scheduling intervals, independent of any iteration and – as explained – independent of any information regarding the previous processing of the subtasks.

3. Suitability for application of the protected method in the case of non-pre-installed software

Even in the event that the Run:ai software is not pre-installed, the suitability of the contested embodiment for applying the protected method cannot be established. It has not been argued that the hardware (mere DGX system products without Run:ai software) – possibly equipped with other software – is suitable for applying the method according to claim 1 of the patent at issue.

C Counterclaim for declaration of invalidity

182 There is no need to rule on the counterclaim for revocation.

183 The defendants have made the decision on the counterclaim for annulment contingent upon the infringement claim being successful in all other respects. As the latter is not the case, no further decision on the counterclaim is required.

184 The panel considers it procedurally permissible for the defendant to bring a counterclaim for revocation in response to an infringement action and to make this subject to the condition subsequent that the infringement action is unsuccessful. If this condition is met because the patent at issue is not infringed, regardless of the legal validity of the patent at issue, no further decision is required on the counterclaim for revocation.

185 It can be left open whether this constitutes an amendment to the claim pursuant to Rule 263.3 of the RoP, a partial withdrawal of the claim pursuant to Rule 265 of the RoP, or even a form of settlement within the meaning of Rule 360 of the RoP, where the defendant only declares during the ongoing proceedings that the decision on the counterclaim is to be dependent on the outcome of the infringement action (also leaving this open: Mannheim local division, decision of 5 December 2025, UPC_CFI_414/2024 – Centripetal v Keysight; in favour of Rule 263.3 of the RoP: Munich local division (Panel 1), decision of 13 January 2026, UPC_CFI_628/2024 – Emboline v. Aorticlab). If the defendant has made the counterclaim for revocation dependent on the outcome of the infringement proceedings, the claimant generally has no interest in the counterclaim being decided even if the infringement action is otherwise unsuccessful. It is the defendant and counter-claimant who articulates the legal interest in a counter-claim for annulment. Where the defendant, as in the present case, makes the decision on the counter-claim dependent on the outcome of the infringement proceedings, this interest lies solely in the defence against the infringement claim (see Rule 25.1 of the RoP). This is because the counterclaim for declaration of invalidity is intrinsically linked to the infringement claim and is generally the direct consequence of the infringement claim brought by the claimant (Court of Appeal, Order of 20 June 2025, UPC_CoA_393/2025 – Emboline v Aorticlab). If the defence against the infringement action is otherwise successful, there is therefore no longer any interest in a decision on the counterclaim. Nor can the claimant have such an interest outside the scope of Article 105a EPC, as they run the risk of unintentionally losing the patent should the counterclaim be successful. However, even the possible dismissal of the counterclaim does not, as a rule, constitute grounds for refusing the defendant’s conditional counterclaim. In so far as the claimants expressed an interest at the oral hearing in wishing to communicate a decision in their favour on the counterclaim to other potential opponents in order to encourage out-of-court discussions, this cannot constitute a legal interest on the part of the claimants. This is because the court’s decision in such a case is limited to dismissing the counterclaim. The validity of a patent is not positively established; the counterclaim is not directed at such a confirmation of validity and cannot be demanded by the claimant and counter-defendant. Furthermore, the dismissal of the counterclaim is not binding on third parties or on other courts. The mere possibility that a potential opponent of the claimants might refer to the dismissal of the counterclaim in out-of-court discussions in his

Although the fact that these considerations are taken into account and that this leads to an out-of-court settlement, thereby avoiding further legal proceedings and the associated costs, may appear advantageous to the claimants, it does not constitute a legal interest.

186 Whether a decision must be made on the costs of the counterclaim for annulment if no decision is reached on the counterclaim has not yet been clarified by the highest court (see Mannheim local division, decision of 5 December 2025, UPC_CFI_414/2024 – Centripetal v Keysight; Munich local division (Panel 1), decision of 13 January 2026, UPC_CFI_628/2024 – Emboline v. Aorticlab). However, this does not require a final decision even in the event of a dispute, as the parties have agreed on a cost allocation that takes precedence over any court decision on costs. Accordingly, the parties shall each bear the costs incurred by them in connection with the infringement action and the counterclaim for annulment.

DECISION

I. The infringement action is dismissed.

II. The parties shall each bear their own costs incurred in connection with the infringement action and the counterclaim for annulment. No reimbursement of costs shall take place.

III. The value in dispute of the infringement action is set at EUR 1,500,000.00. The value in dispute of the counterclaim for annulment is set at EUR 2,250,000.00.

Dr D. Voß (presiding judge)	
Dr G. Werner (Legally qualified judge)	
A. Kupecz (Legally qualified judge)	

A. Dumont (Technically qualified judge)	
For the Deputy-Registrar	

INFORMATION ON APPEALS

Any party whose applications have been rejected in whole or in part may appeal against this decision to the Court of Appeal within two months of the decision being served (Art. 73(1) UPC Agreement, Rule 220.1(a), 224.1
(a) of the RoP).

INFORMATION ON ENFORCEMENT

The decision is not enforceable.

The decision was pronounced in open court on 11 March 2026.

Note
This document is the redacted version of the decision, from which confidential information has been removed. It is valid without the signatures of the judges involved and the representative of the Deputy-Registrar.

Daniel Voß

Digitally signed by Daniel Voß
Date: 11 March 2026 11:40:51 +01'00'