

Convocatoria de Ayudas Beatriz Galindo

Research at the Department of Computer Systems and Computation (DSIC) of the Universidad Politécnica de Valencia (UPV) covers, among others, all areas of computer science and engineering related to algorithms and software. Most of the 151 academics in the department have double affiliation and contribute to and/or lead research groups attached to UPV's research institutes, such as, e.g., the Valencian Research Institute of Artificial Intelligence (VRAIN). The number of publications, funding, technology transfer contracts and other indicators of the aggregate research and transfer activity of DSIC are among those with the highest impact at UPV.

DSIC/UPV research needs in CSE and its key synergies with high performance computing and deep learning. DSIC will greatly benefit from expanding its current research portfolio and capabilities in **Computational Science and Engineering (CSE)** – also known as (a.k.a.) Scientific Computing (SC), and its key synergies with **High Performance Computing (HPC)** – a.k.a. supercomputing, and efficient **Artificial Intelligence (AI)**. CSE is a **highly interdisciplinary** discipline on its own which lies at the intersection of computing (including supercomputing, data science and AI), applied mathematics (including functional and numerical analysis), and most branches of science/engineering (e.g., biology, chemistry, fluid dynamics, solid mechanics, geophysics, climate, etc.). It integrates techniques and methodologies from these fields to develop novel **scalable data-driven numerical algorithms** and **HPC scientific software** for the realistic simulation of complex (e.g. physical or biomedical) phenomena in high-end computing systems (e.g., HPC clouds or state-of-the-art exascale supercomputers).

As in many other disciplines, **big data and deep learning** (DL) have kick-in strongly in CSE. However, as opposed to other disciplines in which breakthroughs in DL have yielded outstanding results (particularly those with abundant high-quality data, such as natural language processing, medical diagnosis or financial forecasting), many CSE applications are characterized by the scarcity of data, which is often indirect, noisy, and expensive to acquire. This caveat can be mitigated *by augmenting DL models with the mathematical models that govern the underlying physical systems*, a methodology referred to as **Scientific Machine Learning (SciML)**. In a nutshell, physical knowledge in the form of mathematical physics models along with data guide the optimization process of the DL system, unlocking the potential of DL for scientific discovery, even with limited amounts of data. The hybridization of more detailed physical models, state-of-the-art numerical methods (e.g., multiscale finite element methods) and DL methodologies (e.g., neural network surrogates) offers striking opportunities to reduce the computational demands and software development efforts while increasing accuracy of state-of-the-art CSE computational models. Through the development and exploration of novel SciML methodologies and techniques, the DSIC is also seeking to reinforce its current partnership and collaboration avenues with the researchers of UPV's Valencian Research Institute of Artificial Intelligence (VRain), which is one of the top players in AI at the national level.

Scientific, technology transfer and societal impact. CSE, through its synergy with HPC and DL, has been established nowadays as the third pillar of science (along with

theoretical/experimental sciences), and it is expected to have an even more prominent role for scientific discovery in the forthcoming years. It bears the potential to revolutionize decision-making across science, technology and society. Grand-challenge applications areas of CSE include geophysics, clean energy/nuclear fusion, personalized health and medicine, and climate/environment, to name but a few. All the grand-challenges in CSE are directly connected with **United Nations (UN) sustainable development goals (SDG)**. Therefore, **CSE has a strong technology transfer and societal impact**. Its potential for enhancing business competitiveness is recognized by the European Commission with currently on-going initiatives such as, e.g., FFplus¹. Such initiatives seek to overcome barriers for small and medium enterprises (SME) by fostering the use of CSE, HPC and DL in order to develop new and optimize existing products, services and/or business processes, with the main aim to strengthen the global competitiveness of European industry. Through its wide adoption in industrial sectors, it also has a capital role in meeting UN's SDG. A paradigmatic example is the aeronautics/aerospace sector and its use of, e.g., advanced shape/topology optimization, and additive manufacturing simulation tools. Such new promising virtual methodologies for aircraft design are capable of reducing fuel consumption by lightening the weights and optimizing the aircraft structures. The possibilities resulting from CSE, HPC, and DL in this context start with a more precise and efficient design of complex work pieces, continue with speeding-up the time-to-market, and conclude with the cost-saving of materials and time savings which could be spent for further R&D. Breakthroughs in algorithms, software and hardware innovations (e.g., a specialized hardware/software co-design for neural network training and inference) for CSE, HPC and DL can also help in reducing carbon emissions, as indeed a non-negligible part of these emissions in developed countries nowadays comes from energy production required to feed computing systems in data centers to train and run these models. CSE, HPC and DL also contribute to reducing carbon footprint by improving the industry energy efficiency, e.g., by discovering more efficient manufacturing and design processes, requiring less amount of raw materials, resources and prototyping.

Inverse PDE problems are cornerstone, highly computationally demanding problems in CSE. Among the array of different problems that CSE addresses, the *computational solution of inverse multi-physics, multi-scale mechanistic systems of Partial Differential Equations (PDEs) in high-end computing systems* is an essential tool in the aforementioned application areas, and thus is *of special relevance for the research at DSIC/UPV*. The ever-increasing demand for resolution and accuracy in real-world CSE applications renders the development of breakthrough inventions in numerical methods, DL methodologies, hardware and software innovations necessary to unravel all the potential of CSE while keeping computational and energy demands within reasonable margins.

In a nutshell, PDEs are continuum models that encode the laws of nature using first physics principles (such as, e.g., the incompressible Navier-Stokes, compressible Euler, or Maxwell equations) expressed in terms of integrals and partial derivatives of multivariate fields defined on a continuous (and potentially geometrically complex) domain. Solving such equations in the context of real-world applications is a daunting task, out of reach by analytical methods due to the lack of closed form solutions in the general case. However, one can recast the continuous problem to approximate it computationally. Mathematically-certified grid-based discretization methods (e.g., finite element methods) are used to transform the continuum problem into a discrete system that can be solved computationally, and fast scalable solvers (e.g., multigrid or multilevel domain decomposition solvers) for the discrete equations resulting from discretization can be leveraged to tackle the solution of the discrete problem on

¹ Website of FFPlus initiative available at <https://www.ffplus-project.eu/>.

massively parallel high-end computing systems.

Unfortunately, in many applications, the input parameters to the PDE might be unknown and not directly observable via experiments, **rendering an inverse problem**. In this setting, the model parameters are not fully known, but one can obtain some observations, typically noisy and/or partial, of the model state and parameters. Inverse problem solvers combine the partially known model and the observations to infer the information which is missing to complete the model. This process is a.k.a. **data assimilation**. Inverse problems are typically modelled using PDE-constrained optimization. The de-facto standard method to solve these problems is **the adjoint method, which is a formidably expensive algorithm** (as it requires the imposition of the (discretized) PDE as a hard constraint during the optimization process). To reduce computational demands (among other striking benefits), **big data and DL techniques (e.g., when used as surrogate models) have emerged as especially appealing alternatives**. Unlike classical methods, neural networks are parameterised manifolds that enjoy nonlinear adaptive approximation power, allowing them to handle complex multiscale problems more effectively. They also offer seamless data integration for PDE-constrained inverse problems and efficient minimisation by leveraging non-local parameterisations to deliver faster alternatives to adjoint methods.

Standard simulation pipelines fall short in meeting the simulation and computational demands of state-of-the-art CSE application problems. Nowadays, one routinely encounters transient, highly nonlinear PDE models that feature multiple scales in time and space on complex geometrical domains. Besides, models involve the coupling of multiple physics in the bulk of the domain and/or complex geometrical interfaces among different regions of the domain. To make things worse, in some application problems the geometry of the domain grows in time and/or the geometrical interface among different physics evolves in time, and is unknown a-priori, as in free boundary problems. Besides, physical parameters of the PDE might be heterogeneous and discontinuous across high-dimensional, unknown, geometrically intricate interfaces.

Standard simulation pipelines (e.g., those leveraged in industrial scenarios, and available in commercial software) based on unstructured, body-fitted mesh generation (e.g., gmsh) conforming mesh adaptation, and graph partitioning (e.g., METIS/ParMETIS) for dynamic load balancing, space-time FEM discretization using the method of lines (MOL), and non-linear solver algorithms (e.g., Newton-Raphson linearization) combined with scalable preconditioned iterative sparse linear solvers (Geometric MultiGrid-GMG or multilevel domain decomposition) at each nonlinear solver step, all wrapped around a further external optimization loop that leverages the adjoint method, **fall short in tackling the aforementioned modelling challenges in state-of-the-art CSE PDE mathematical models**. The high computational demands of such models require the efficient exploitation of the most powerful computers across the globe, and the integration of methodologies and techniques from supercomputing. Without investment in algorithmic inventions and supercomputing, the efficient/scalable exploitation of the vast amount of parallelism available in state-of-the-art exascale computers is just not possible.

We find ourselves in the Exascale computing era i.e., large-scale distributed-memory supercomputers able to deliver $O(10^{18})$ floating point operations per second at its peak. Two such supercomputers are Frontier and El Capitan at Oak Ridge and Lawrence Livermore USA National Labs, resp. Such tremendous computing power is achieved with heterogeneous computing architectures that combine multi-core CPUs and many-core GPUs within each compute node along with breakthrough network and computer architecture innovations to keep energy constraints within reasonable margins. **To efficiently exploit the massive**

amount of hardware parallelism readily available at the different computing system levels (in the order of several millions of computational cores) **is a formidable task for CSE**, and it requires significant **advances in scalable algorithms and software and hardware innovations**. The synergies among more realistic physical models, state-of-the-art numerical methods for PDEs (e.g., finite element methods) and DL tools (e.g., neural network surrogates) offer capital opportunities to reduce the computational demands and software development efforts while increasing accuracy of state-of-the-art CSE computational models. This challenge is being tackled worldwide by initiatives such as the so-called Trillium Parameter Consortium.² In connection with the hardware and software innovations, UPV, through the Department of Computing Engineering (DISCA), is pursuing along with other top institutions in high-end computing such as BSC, the development of computing architectures based on RISC-V architectures, and software/hardware co-design strategies tailored to accelerate CSE and DL computing. This is being funded by the [DARE](#) Horizon Europe project during a period of 6 years. Strengthening current synergies among DISCA and DSIC are thus also of capital importance for the success of this research project.

A new research line at DSIC/UPV. The proposed research plan provided by the candidate will aim at the development and integration of next generation numerical methods, scientific machine learning methodologies and HPC scientific software able to tackle several urgent bottlenecks/limitations of existing methods, techniques and methodologies that hinder their ability to meet the required modelling, simulation and computational goals. The approach that will be pursued is based on a combination of algorithmic and software inventions at the intersection of CSE, AI and HPC. There is also the goal to implement all these algorithms in a HPC differentiable programming open source scientific software framework by combining source-to-source Automatic Differentiation (AD), CSE and ML software libraries. To this end, the **Julia programming language** and its rich package ecosystem³ constitute a 21st century, open-source, multi-platform, high-level, interactive, high-performance software framework for technical computing, which is especially well suited and will be leveraged to satisfy the goals of the project. Besides, boosted by the excellent capabilities of Julia, its rich package ecosystem, professional tools for development workflow (IDEs, debuggers, profilers, visualization etc.), and interactive educational tools, **the software developed as part of the project will also be well suited for teaching.**

Candidate expectations and contributions to DSIC/UPV. It is expected that the candidate creates its own research group at the interface of CSE, HPC and DL and leads this novel, timely, high-risk/high-gain, and cutting-edge research line within DSIC/UPV. Being a highly visible subject, and having a primary focus on open-source software development with scientific reproducibility and community building in mind, the project is expected to bring several new collaborations to DSIC/UPV. The candidate will supervise postdocs, PhDs, and graduate and undergraduate students (via short research projects) in high performance scientific computing and scientific machine learning, sought-after skills vital to our nation's economy. The results will be disseminated to the scientific community in the form of high impact publications in world-class research journals, presentations at international conferences, scientific workshops, either to specialized audiences in CSE or via outreach activities to the general public, and patented products. Besides, the new algorithmic inventions resulting from the project will be implemented in generic open-source software packages that will be available to the scientific community and industry in public software repositories (e.g., GitHub). The current state-of-the-art in this research topic suffers from a

² Website of the TPC available at <https://tpc.dev/>.

³ See, e.g., <https://github.com/JuliaPDE/SurveyofPDEPackages> or <https://sciml.ai/>.

concerning lack of reproducibility of computational results and high-quality software. This results in a poor understanding of the relative merits of the different approaches and non-standardised benchmarking. Apart from increased R&D capability, high-quality open-source software enables reproducibility in computational science, and creates new opportunities for technology and knowledge transfer to industry.

The scientific outcomes will significantly expand UPV's knowledge of state-of-the-art computational simulation techniques and reinforce the university position as a leader in open-source scientific software. The line is highly aligned with the objectives, principles and mission areas of the State Plan for Scientific and Technical Research and Innovation 2024-2027, as well with those of the Horizon Europe framework by the European Commission (i.e., excellent science, openness, tackling global society challenges such as health, climate change, and clean energy - UN's SDG). It is also aligned with international on-going R&D initiatives, such as the IN-DEEP European Doctoral Network⁴, and the goals of top industry players in CSE, HPC and ML, such as, e.g. NVIDIA⁵ or Microsoft⁶. Exploiting such alignment, the candidate will attract national and international funding, engage in technology transfer initiatives, while creating future job opportunities for supervised research students and postdocs.

Given the intrinsic multidisciplinary nature of the proposed research, this endeavour will require expertise from different fundamental research areas that already exist within the DSIC department as well as other departments and institutes of the UPV such as the Department of Computer Engineering (DISCA), the Institute for Telecommunications and Multimedia Applications (ITEAM), or the Department of Applied Mathematics. The candidate should also establish partnerships with the UPV-I2MB research institute in mechanical and biomechanical engineering to, e.g., explore their technology and knowledge transfer avenues for the newly developed computational tools and HPC software. Besides, it also expected that the candidate brings its current partnerships or explores new ones with top R&D and technology transfer players at the national (e.g., CIMNE/UPC, CIMNE Tecnología, BSC-CASE, UPV/EHU, BCAM, ICMAB-CSIC, CIEMAT-LNF, BSC) and international level (e.g., Monash University, Curtin University, University of Oslo, Simula, Oxford University, TU Delft, VU Amsterdam).

⁴ Network's website available at <https://www.in-deep.science/>.

⁵ See NVIDIA SCiML Modulus framework (<https://developer.nvidia.com/modulus>).

⁶ See Microsoft's AI4Science division

(<https://www.microsoft.com/en-us/research/lab/microsoft-research-ai-for-science/>)

La investigación en el Departamento de Sistemas Informáticos y Computación (DSIC) de la Universidad Politécnica de Valencia (UPV) abarca, entre otras, todas las áreas de la informática y la ingeniería relacionadas con los algoritmos y el software. La mayoría de los 151 académicos del departamento tienen doble adscripción y contribuyen y/o lideran grupos de investigación adscritos a institutos de investigación de la UPV, como, por ejemplo, el Instituto Valenciano de Investigación en Inteligencia Artificial (VRAIN). El número de publicaciones, la financiación, los contratos de transferencia de tecnología y otros indicadores de la actividad agregada de investigación y transferencia del DSIC se encuentran entre los de mayor impacto de la UPV.

Necesidades de investigación del DSIC/UPV en CSE y sus sinergias clave con la computación de alto rendimiento y el aprendizaje profundo. El DSIC se beneficiará enormemente de la ampliación de su actual cartera de investigación y capacidades en **Ciencia e Ingeniería Computacional (CSE)** - también conocida como Computación Científica (SC), y sus sinergias clave con la **Computación de Altas Prestaciones (HPC)** - también conocida como supercomputación, y la **Inteligencia Artificial (IA)** eficiente. La CSE es una disciplina **altamente interdisciplinar** que se encuentra en la intersección de la informática (incluida la supercomputación, la ciencia de datos y la IA), las matemáticas aplicadas (incluido el análisis funcional y numérico) y la mayoría de las ramas de la ciencia y la ingeniería (por ejemplo, biología, química, dinámica de fluidos, mecánica de sólidos, geofísica, clima, etc.). Integra técnicas y metodologías de estos campos para desarrollar nuevos **algoritmos numéricos escalables basados en datos y software científico HPC** para la simulación realista de fenómenos complejos (por ejemplo, físicos o biomédicos) en sistemas informáticos de alta gama (por ejemplo, nubes HPC o superordenadores exaescala de última generación).

Como en muchas otras disciplinas, el **“big data” y el aprendizaje profundo (deep learning, DL)** han irrumpido con fuerza en la CSE. Sin embargo, a diferencia de otras disciplinas en las que los avances en DL han dado resultados sobresalientes (en particular aquellas con abundantes datos de alta calidad, como el procesamiento del lenguaje natural, el diagnóstico médico o la previsión financiera), muchas aplicaciones de CSE se caracterizan por la escasez de datos, que a menudo son indirectos, ruidosos y caros de adquirir. Este problema puede mitigarse aumentando los modelos de DL con los modelos matemáticos que gobiernan los sistemas físicos subyacentes, una metodología denominada **aprendizaje automático científico (Scientific Machine Learning, SciML)**. En pocas palabras, el conocimiento físico en forma de modelos de física matemática junto con los datos guía el proceso de optimización del sistema de DL, liberando su potencial para el descubrimiento científico, incluso con cantidades limitadas de datos. La hibridación de modelos físicos más detallados, métodos numéricos de vanguardia (por ejemplo, métodos de elementos finitos multiescala) y metodologías de DL (por ejemplo, redes neuronales surrogadas) ofrece sorprendentes oportunidades para reducir las demandas computacionales y los esfuerzos de desarrollo de software, al tiempo que se incrementa la precisión de los modelos computacionales CSE de vanguardia. A través del desarrollo y la exploración de nuevas metodologías y técnicas SciML, el DSIC también pretende reforzar su actual asociación y vías de colaboración con los investigadores del Instituto Valenciano de Investigación en Inteligencia Artificial (VRAIN) de la UPV, que es uno de los principales actores en IA a nivel nacional.

Impacto científico, tecnológico y social. La CSE, a través de su sinergia con la HPC y el DL, se ha establecido hoy en día como el tercer pilar de la ciencia (junto con las ciencias teóricas/experimentales), y se espera que tenga un papel aún más destacado en el descubrimiento científico en los próximos años. Tiene el potencial de revolucionar la toma de

decisiones en la ciencia, la tecnología y la sociedad. Entre las áreas de aplicación de la CSE que plantean grandes retos figuran la geofísica, la energía limpia y la fusión nuclear, la salud y la medicina personalizadas y el clima y el medio ambiente, por citar sólo algunas. Todos los grandes retos de la CSE están directamente relacionados con los **objetivos de desarrollo sostenible (ODS) de las Naciones Unidas**. Por lo tanto, **la CSE tiene un fuerte impacto en la transferencia de tecnología y en la sociedad**. Su potencial para mejorar la competitividad de las empresas está reconocido por la Comisión Europea con iniciativas actualmente en curso como, por ejemplo, FFplus⁷. Dichas iniciativas pretenden superar las barreras a las que se enfrentan las pequeñas y medianas empresas (PYME) fomentando el uso de la CSE, la HPC y el DL para desarrollar nuevos productos, servicios y/o procesos empresariales y optimizar los ya existentes, con el objetivo principal de reforzar la competitividad global de la industria europea. Gracias a su amplia adopción en los sectores industriales, también desempeña un papel capital en el cumplimiento de los ODS de las Naciones Unidas. Un ejemplo paradigmático es el sector aeronáutico/aeroespacial y su uso, por ejemplo, de herramientas avanzadas de optimización de formas/topologías y de simulación de fabricación aditiva. Estas nuevas y prometedoras metodologías virtuales para el diseño de aeronaves son capaces de reducir el consumo de combustible al aligerar los pesos y optimizar las estructuras de las aeronaves. En este contexto, las posibilidades derivadas de la CSE, la HPC y el DL comienzan con un diseño más preciso y eficiente de piezas complejas, continúan con la aceleración de los plazos de comercialización y concluyen con el ahorro de costes de materiales y de tiempo, que podría dedicarse a más I+D. Los avances en algoritmos e innovaciones de software y hardware (por ejemplo, un codiseño especializado de hardware/software para el entrenamiento y la inferencia de redes neuronales) para CSE, HPC y DL también pueden ayudar a reducir las emisiones de carbono, ya que, de hecho, una parte no desdeñable de estas emisiones en los países desarrollados procede hoy en día de la producción de energía necesaria para alimentar los sistemas informáticos de los centros de datos para entrenar y ejecutar estos modelos. La CSE, la HPC y el DL también contribuyen a reducir la huella de carbono mejorando la eficiencia energética de la industria, por ejemplo, descubriendo procesos de fabricación y diseño más eficientes, que requieren menos cantidad de materias primas, recursos y prototipos.

Los problemas de EDP inversas son la piedra angular en numerosos problemas de alta exigencia computacional de la CSE. Entre los diferentes problemas que aborda la CSE, *la solución computacional de sistemas mecánicos inversos multifísicos y multiescala de Ecuaciones Diferenciales Parciales (EDP) en sistemas de computación de alta gama* es una herramienta esencial en las áreas de aplicación mencionadas, y por lo tanto es ***de especial relevancia para la investigación en el DSIC/UPV***. La demanda cada vez mayor de resolución y precisión en las aplicaciones CSE del mundo real hace necesario el desarrollo de invenciones revolucionarias en métodos numéricos, metodologías DL e innovaciones de hardware y software para desentrañar todo el potencial de CSE manteniendo las demandas computacionales y energéticas dentro de unos márgenes razonables.

En pocas palabras, las EDP son modelos continuos que codifican las leyes de la naturaleza utilizando los primeros principios de la física (como las ecuaciones no comprimibles de Navier-Stokes, Euler o Maxwell) expresadas en términos de integrales y derivadas parciales de campos multivariantes definidos en un dominio continuo. Resolver estas ecuaciones en aplicaciones reales es una tarea de enormes proporciones, fuera del alcance de los métodos analíticos debido a la falta de soluciones en el caso general. Sin embargo, se puede refundir el problema continuo para aproximarlos computacionalmente utilizando métodos de

⁷ La página web de la iniciativa FFPlus está disponible en <https://www.ffplus-project.eu/>.

discretización basados en mallas con certificación matemática (por ejemplo, métodos de elementos finitos), y abordando entonces la solución del problema discreto en sistemas informáticos de alto nivel masivamente paralelos pueden aprovecharse algoritmos de resolución rápidos y escalables (por ejemplo, métodos multimalla o de descomposición de dominios multinivel) para las ecuaciones discretas resultantes de la discretización.

Por desgracia, en muchas aplicaciones, los parámetros de entrada de la EDP pueden ser desconocidos y no observables directamente mediante experimentos, lo que **plantea un problema inverso**. En este caso, los parámetros del modelo no se conocen por completo, pero se pueden obtener algunas observaciones, normalmente ruidosas y/o parciales, del estado y los parámetros del modelo. Los resolutores de problemas inversos combinan el modelo parcialmente conocido y las observaciones para deducir la información que falta para completar el modelo. Este proceso se conoce como **asimilación de datos**. Los problemas inversos suelen modelizarse mediante optimización con restricciones PDE. El método estándar de facto para resolver estos problemas es el **método adjunto, que es un algoritmo formidablemente caro** (ya que requiere la imposición de la EDP discretizada como una restricción dura durante el proceso de optimización). Para reducir las exigencias computacionales, **el big data y las técnicas de DL (por ejemplo, cuando se utilizan como modelos sustitutos) han surgido como alternativas especialmente atractivas**. A diferencia de los métodos clásicos, las redes neuronales son modelos parametrizados que disfrutan de un poder de aproximación adaptativo no lineal, lo que les permite manejar problemas complejos multiescala con mayor eficacia. También ofrecen una integración de datos perfecta para los problemas inversos con restricciones PDE y una minimización eficaz al aprovechar las parametrizaciones no locales para ofrecer alternativas más rápidas a los métodos adjuntos.

Los algoritmos de simulación estándar no satisfacen las demandas de simulación y cálculo de los problemas de aplicación de la CSE más avanzados. En la actualidad, es habitual encontrarse con modelos de EDP transitorios y altamente no lineales que presentan múltiples escalas en el tiempo y el espacio en dominios geométricos complejos. Además, los modelos implican el acoplamiento de múltiples físicas en el volumen del dominio y/o interfaces geométricas complejas entre diferentes regiones del dominio. Para empeorar las cosas, en algunos problemas la geometría del dominio crece en el tiempo y/o la interfaz geométrica entre las diferentes físicas evoluciona en el tiempo, y es desconocida a-priori. Además, los parámetros físicos de la EDP pueden ser heterogéneos y discontinuos a través de interfaces de alta dimensión, desconocidas y geoméricamente intrincadas.

Los algoritmos de simulación estándar (por ejemplo, utilizados en escenarios industriales y disponibles en software comercial) se basan en la generación de mallas no estructuradas, la adaptación de mallas y la partición de grafos (p.e., METIS/ParMETIS) para el equilibrio dinámico de la carga, la discretización espacio-temporal mediante el método de las líneas (MOL) y los algoritmos de resolución no lineal (p.e., linealización de Newton-Raphson) combinados con solvers lineales dispersos iterativos preconditionados escalables en cada paso del solucionador no lineal (e.g., métodos multimalla geométricos o métodos multiescala de descomposición del dominio), todo ello envuelto en un bucle de optimización externo adicional que aprovecha el método adjunto, **se quedan cortos al abordar los retos de modelización mencionados en los modelos matemáticos de EDP de CSE de última generación**. Las elevadas exigencias computacionales de estos modelos requieren el aprovechamiento eficiente de los ordenadores más potentes del mundo y la integración de metodologías y técnicas de supercomputación. Sin inversión en invenciones algorítmicas y supercomputación, el aprovechamiento eficiente/escalable de la enorme cantidad de paralelismo disponible en los ordenadores de exaescala de última generación no es posible.

Nos encontramos en la era de la computación a Exaescala, es decir, superordenadores de memoria distribuida a gran escala capaces de realizar $O(10^{18})$ operaciones en coma flotante por segundo de pico. Dos de estos superordenadores son Frontier y El Capitán, situados en los laboratorios nacionales de EEUU de Oak Ridge y Lawrence Livermore, respectivamente. Esta enorme potencia de cálculo se consigue con arquitecturas de computación heterogénea que combinan CPU y GPU en cada nodo de cálculo, junto con innovaciones revolucionarias en la arquitectura de redes y ordenadores para mantener las limitaciones energéticas dentro de unos márgenes razonables. **Aprovechar eficientemente la enorme cantidad de paralelismo de hardware** disponible en los distintos niveles del sistema informático (del orden de varios millones de núcleos computacionales) **es una tarea formidable para la CSE**, y requiere **avances significativos en algoritmos escalables e innovaciones de software y hardware**. Las sinergias entre modelos físicos más realistas, métodos numéricos de vanguardia para las EDP (por ejemplo, métodos de elementos finitos) y herramientas de DL (por ejemplo, redes neuronales surrogadas) ofrecen oportunidades capitales para reducir las demandas computacionales y los esfuerzos de desarrollo de software, aumentando al mismo tiempo la precisión de los modelos computacionales de vanguardia de la CSE. Este reto está siendo abordado en el mundo por iniciativas como Trillium Parameter Consortium⁸. En conexión con las innovaciones hardware y software, la UPV, a través del Departamento de Ingeniería Informática (DISCA), persigue junto con otras instituciones punteras en computación de alto rendimiento, como el BSC, el desarrollo de arquitecturas de computación basadas en arquitecturas RISC-V, y estrategias de co-diseño software/hardware a medida para acelerar la computación CSE y DL. Este proyecto está financiado por el proyecto DARE Horizon Europe durante un periodo de 6 años. Reforzar las sinergias actuales entre DISCA y DSIC es, por tanto, también de capital importancia para el éxito de este proyecto de investigación.

Una nueva línea de investigación en DSIC/UPV. El plan de investigación propuesto por el candidato tendrá como objetivo el desarrollo y la integración de métodos numéricos de nueva generación, metodologías científicas de aprendizaje automático y software científico HPC capaz de hacer frente a varios cuellos de botella/limitaciones urgentes de los métodos, técnicas y metodologías existentes que dificultan su capacidad para cumplir con los objetivos de modelización, simulación y computación requeridos. El enfoque que se seguirá se basa en una combinación de invenciones algorítmicas y de software en la intersección de la CSE, la IA y la HPC. También existe el objetivo de implementar todos estos algoritmos en un marco de software científico de código abierto de programación diferenciable HPC mediante la combinación de bibliotecas de software de fuente a fuente de Diferenciación Automática (AD), CSE y ML. Para ello, el **lenguaje de programación Julia** y su rico ecosistema⁹ de paquetes constituyen un marco de software del siglo XXI, de código abierto, multiplataforma, de alto nivel, interactivo y de alto rendimiento para la computación técnica, que es especialmente adecuado y se aprovechará para satisfacer los objetivos del proyecto. Además, impulsado por las excelentes capacidades de Julia, su rico ecosistema de paquetes, herramientas profesionales para el flujo de trabajo de desarrollo (IDEs, depuradores, perfiladores, visualización, etc.), y herramientas educativas interactivas, **el software desarrollado como parte del proyecto también será muy adecuado para la enseñanza**.

Expectativas del candidato y contribuciones al DSIC/UPV. Se espera que el candidato cree su propio grupo de investigación en la línea de CSE, HPC y DL y lidere esta línea de investigación novedosa, oportuna, de alto riesgo/alto beneficio, y de vanguardia dentro de DSIC/UPV. Al tratarse de un tema de gran visibilidad y centrarse principalmente en el

⁸ La página web de TPC está disponible en <https://tpc.dev/>.

⁹ Ver, p.e., <https://github.com/JuliaPDE/SurveyofPDEPackages> o <https://sciml.ai/>.

desarrollo de software de código abierto teniendo en cuenta la reproducibilidad científica y la creación de comunidades, se espera que el proyecto aporte varias colaboraciones nuevas al DSIC/UPV. El candidato supervisará a doctores y estudiantes de grado y postgrado (a través de proyectos de investigación cortos) en computación científica de altas prestaciones y aprendizaje automático científico, habilidades muy solicitadas y vitales para la economía de nuestro país. Los resultados se difundirán a la comunidad científica en forma de publicaciones de alto impacto en revistas de investigación de categoría mundial, presentaciones en congresos internacionales, talleres científicos, ya sea a audiencias especializadas en CSE o mediante actividades de divulgación al público en general, y productos patentados. Además, las nuevas invenciones algorítmicas resultantes del proyecto se implementarán en paquetes genéricos de software de código abierto que estarán disponibles para la comunidad científica y la industria en repositorios públicos de software (GitHub). El estado actual de este tema de investigación adolece de una preocupante falta de reproducibilidad de los resultados y de software de alta calidad. Esto da lugar a una escasa comprensión de los méritos relativos de los distintos enfoques y a una evaluación comparativa no estandarizada. Además de aumentar la capacidad de I+D, el software de código abierto de alta calidad crea nuevas oportunidades para la transferencia de tecnología y conocimientos a la industria.

Los resultados científicos ampliarán significativamente el conocimiento de la UPV sobre las técnicas de simulación computacional más avanzadas y reforzarán la posición de la UPV como líder en software científico de código abierto. La investigación está muy alineada con los objetivos, principios y áreas del Plan Estatal de Investigación Científica y Técnica y de Innovación 2024-2027, así como con los del Horizonte Europa de la Comisión Europea (ciencia excelente, apertura, abordar los retos de la sociedad global como la salud, el cambio climático y la energía limpia-ODS de las Naciones Unidas). También está en consonancia con iniciativas internacionales de I+D en curso, como la Red Europea de Doctorado IN-DEEP¹⁰, y los objetivos de los principales actores de la industria en CSE, HPC y ML, como, por ejemplo, NVIDIA¹¹ o Microsoft¹². Aprovechando esta alineación, el candidato atraerá financiación nacional e internacional, participará en iniciativas de transferencia de tecnología y creará futuras oportunidades de empleo para los estudiantes de investigación y postdoctorales supervisados.

Dada la naturaleza intrínsecamente multidisciplinar de la investigación propuesta, este esfuerzo requerirá la experiencia de diferentes áreas de investigación fundamentales que ya existen dentro del departamento DSIC, así como de otros departamentos e institutos de la UPV como el DISCA, el Instituto de Telecomunicaciones y Aplicaciones Multimedia (ITEAM), o el Departamento de Matemática Aplicada. El candidato también deberá establecer colaboraciones con el instituto de investigación UPV-I2MB en ingeniería mecánica y biomecánica para, por ejemplo, explorar sus vías de transferencia de tecnología y conocimiento para las herramientas computacionales y el software HPC recientemente desarrollados. Además, también se espera que el candidato aporte sus relaciones actuales o explore otros nuevos con los principales actores de I+D y transferencia de tecnología a nivel nacional (por ejemplo, CIMNE/UPC, CIMNE Tecnología, BSC-CASE, UPV/EHU, BCAM, ICMAB-CSIC, CIEMAT-LNF, BSC) e internacional (por ejemplo, Monash University, Curtin University, University of Oslo, Simula, Oxford University, TU Delft, VU Amsterdam).

¹⁰ La página web de la Red está disponible en <https://www.in-deep.science/>.

¹¹ Ver el marco de trabajo NVIDIA SCiML Modulus (<https://developer.nvidia.com/modulus>).

¹² Ver la división Microsoft's AI4Science (<https://www.microsoft.com/en-us/research/lab/microsoft-research-ai-for-science/>).