國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

（編號：NCHC-115-01-F-017）

經費來源：■01 公務　□02 非公務
機密(E)：□是　■否
出國類別：□ A 考察/訪問　□ C 進修/研究　■ F 工作會議/研討會
□ G 推廣佈展　□ H 學術會議

分項計畫名稱：新一代高速運算主機建置與服務

SCA/HPCAsia 2026

18th ADAC Symposium & Workshop

大阪大學參訪與討論

出國報告書

服務單位：　國家實驗研究院國家高速網路與計算中心
出國人姓名職稱：　饒駿頌副研究員

出國地點：　日本
出國日期：　民國 115 年 01 月 25 日至 115 年 02 月 05 日
報告日期：　民國 115 年 03 月 09 日

國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

摘　要

本次出國的主要目的在參加 SCA/HPC Asia 2026 以及 18th ADAC Symposium & Workshop，希望掌握國際高效能運算（HPC）和加速運算技術的最新發展趨勢，也藉此加強與全球 HPC 社群的連結與合作。

行程安排包括：

1. SCA/HPC Asia 2026（1/26 – 1/30）
2. ADAC 18th Symposium（1/29、2/2 – 2/4）

期間，也在 OpenACC User Workshop 分享工作成果「GPU-Accelerated Particle Tracking on Unstructured Meshes Using OpenACC」。在兩個會議之間的周末（1/31），也前往大阪大學拜訪研究合作團隊，完成合作論文「Domain-Adaptive Local Solver for Field Reconstruction in Proton Radiography 並進行投稿。」

本報告主要分享會議中一些亮點，包括：

- OpenACC User Workshop
- AI Scientist
- Mixed-precision Computing
- Fugaku-NEXT 與次世代 HPC-AI 支援中心（HAIRDESC）
- Slurm + Slinky 的跨環境整合
- AMD ROCm 與 TheRock 整合策略

其中，「AI Scientist」和「Mixed-precision Computing」計算是目前比較新興且有趣的議題，另在附錄中也整理了它們的發展回顧。

國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

活動日程表

國別	日期	地點/訪問機構	工作摘要/接待人員
日本	1/25 (日)	台北 → 大阪	路程
	1/26 (一)	大阪 / SCA/HPCAsia 2026	Keynote Speech OpenACC User Workshop Slurm + Slinky for the AI and Supercomputer MxP boosts non-AI tasks with AI-devices Trillion Parameter Consortium ADAC Symposium
	1/27 (二)		
	1/28 (三)		
	1/29 (四)		
	1/30 (五)		
	1/31 (六)	大阪 / 大阪大學	Prof. Yasuhiro Kuramitsu 團隊討論合作研究並完成論文
	2/01 (日)	大阪 → 東京	路程
	2/02 (一)	東京 / ADAC18 Workshop	AMD - Task Force Software Stack Working Group - Applications & Benchmarks
	2/03 (二)		
	2/04 (三)		
	2/05 (四)	東京 → 台北	路程

國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

目　　次

1.目的

為了掌握國際高效能運算（HPC）及加速運算技術的最新發展，本次行程安排出訪 SCA/HPC Asia 2026 以及 18th ADAC Symposium & Workshop。其中，SCA/HPC Asia 2026 是由 SCA（Supercomputing Asia） 與 HPC Asia（High Performance Computing Asia Conference） 共同舉辦的年度會議。

在首週的 SCA/HPC Asia 2026 中，個人特別關注 OpenACC、AI Scientist、Mixed-precision Computing、Fugaku-NEXT 與 HPC-AI 支援中心、以及 Slurm + Slinky 的跨環境整合等議題，並在 OpenACC User Workshop 分享研究成果「GPU-Accelerated Particle Tracking on Unstructured Meshes Using OpenACC 」。在第二週的 ADAC Symposium & Workshop，個人除了參加「Applications & Benchmarks」的工作小組外，也特別關注 AMD ROCm/TheRock 部署等議題。

此外，會議之間的週末，也規劃行程前往大阪大學拜訪 Prof. Yasuhiro Kuramitsu 的研究團隊，針對將 AI 應用於高能雷射實驗中電磁場診斷的合作研究，完成論文修改，並投稿至國際期刊 APL Machine Learning。

整體而言，本次差旅涵蓋 HPC 發展、GPU 加速以及 AI 科學運算等多個議題，蒐集國際前沿經驗，為未來研究量能提升、程式碼 GPU 化及跨機構合作提供具體參考。更多細節將於後續章節中說明。

2.參訪(或進修、研究、實習及會議)紀要

2.1 SCA/HPCAsia 2026

SCA/HPC Asia 2026 是今年由日本主辦、在大阪舉行的聯合會議，結合 SCA 與 HPC Asia 兩大活動共同辦理。

其中，SCA 較聚焦於高效能運算、平行計算、雲端運算與 AI 應用，同時也包含展覽活動。與會者多來自研究機構、政府部門及產業

界的 HPC 專家。討論內容涵蓋最新計算架構、軟體工具、實務應用案例以及效能優化技術等議題。HPC Asia 則是結合學術與產業的研討會，主軸放在 HPC 系統、超級電腦應用、計算科學方法與技術創新。整體方向與 SCA 相近，但更偏重技術交流與學術論文發表。

透過兩個具代表性的會議共同舉辦，匯集來自亞洲及其他地區的高效能運算研究人員、學者、技術專家與產業領袖，使議題交流更為集中，也讓討論產生更大的效果。

2.2 18th ADAC Symposium & Workshop

ADACADAC（Accelerated Data Analytics and Computing Institute）是由全球多個頂尖 HPC 研究中心組成的合作聯盟，主要關注加速運算（特別是 GPU 與混合架構）以及科學與工程應用軟體的可移植性與效能優化。參與單位包括 Oak Ridge National Laboratory、Argonne National Laboratory、RIKEN Center for Computational Science、ETH Zurich / CSCS 等國際重要 HPC 中心，本研究中心亦為成員之一。討論涵蓋科學工程計算的加速化、硬體架構多樣性評估、分散式 AI 與量子計算整合、能源效率與軟體永續性、資料中心維運管理等議題。

本屆會議配合 SCA/HPC Asia 2026 聯合舉行，第一部分的研討會先於大阪舉辦，並開放給所有 SCA/HPC Asia 2026 與會者參加；至於僅限 ADAC 成員參與的 Workshop，則於隔週在東京的產業技術綜合研究所臨海副都心舉行。會議內容除了各主題工作小組討論外，也包含 AMD 廠商的軟體工作小組會議。

本次 ADAC 會議是個人自 2025 年 9 月以來第二次參與。依主管指示，參與的工作小組轉為 Applications & Benchmarks。希望透過持續、長期的參與，深入了解該工作小組的運作模式，並逐步提升本中心在該組織中的參與度與實質貢獻。

2.3 大阪大學參訪與討論

基於本中心電漿團隊與 Osaka University 以及 National Institute for Fusion Science（NIFS）間的科學研究合作關係，個人也利用兩場會議之間的週末空檔拜訪 Prof. Yasuhiro Kuramitsu 團隊，密集討論完成合作論文「Domain-Adaptive Local Solver for Field Reconstruction in Proton Radiography」，並已投稿至 APL Machine Learning。

透過高能雷射所產生的電漿，科學家可以在實驗室中產生與天文現象（例如震波、磁重聯）相似的複雜電磁場結構。不過在利用 proton radiography 量測電磁場時，實際量到的只有質子穿越電漿後，在探測器上的分布影像；這個影像其實是三維電磁場對粒子軌跡積分後的結果，因此是一個高度非線性、且資訊不足的反演問題。再加上真實雷射產生的質子束通常具有寬頻能譜，中心能量也無法固定，使得傳統依賴理論模型或數值模擬的反推方法很容易產生誤差。

在我們的工作中並沒有採用傳統「影像對影像」的深度學習架構，而是將問題拆解成「局部求解器」的概念。利用大量隨機均勻磁場中的單粒子運動軌跡來訓練神經網路，讓模型學習最基本的洛倫茲力物理關係，而不是記住某種特定的磁場拓樸。實際推論時，再把每一顆質子的量測資訊輸入神經網路，最後把所有局部預測整合起來，重建成完整的磁場分布圖。

為了避免能譜展寬造成模型外插失準，我們也加入 OOD（out-of-distribution）過濾機制，只允許落在訓練能量範圍內的粒子參與重建，相當於在前端替模型做第一道把關。即使面對模型從未見過的複雜磁場結構，重建結果仍可達到相關係數 R 大於 0.99、誤差低於 3% 的高準確度。此外，我們提出「In-Distribution fraction」作為事前即可計算的不確定度指標，並進一步預估重建的空間覆蓋率與可靠度。相關技術與實驗設計細節，請參閱附錄三中送審論文初稿。

3.心得及建議

3.1 OpenACC User Workshop

本工作坊的目的，是匯集來自國家實驗室、大學、產業界與研究機構的 OpenACC 使用者，交流 OpenACC 程式模型在各科學領域的實務經驗。活動內容包含使用者技術報告與社群討論，涵蓋不同研究領域中的應用案例，以及未來標準發展方向。整體重點聚焦在跨平台效能可移植性的技術方法，以及科學軟體長期演進的相關議題。

OpenACC 是一種指令式（directive-based）、強調效能可移植性（performance portability）的平行程式設計模型，支援多種類型的加速器，並可與 C、C++、Fortran 相容。其核心設計理念，是協助開發者將既有程式碼從主機（host）有效率地移植到高效能運算加速器（如 GPU），降低平行化與加速過程中的開發成本。某種程度上，OpenACC 也可視為對標準程式語言的補充機制，透過指令方式補強語言本身在平行運算支援上的不足。隨著 ISO 標準（如 C++ 與 Fortran）逐步強化平行運算能力，部分由 OpenACC 推動的概念，也正在被納入官方標準模型的發展方向。

本次議程涵蓋規格更新、編譯器進展（如 GCC、NVIDIA、HPE）、多項科學應用實作案例，以及未來標準演進的討論，重點包括：

- 規格與編譯器發展
- 大氣與氣候模型的 GPU 加速
- 計算科學與物理模擬應用
- 科學軟體實務案例分享

在工作坊主辦單位透過本中心主管的邀請下，我們也利用此次參加 SCA/HPC Asia 2026 的機會，分享本中心 CFD 團隊在非結構化網格粒子追蹤程式上的 GPU 加速實作經驗。過程中也獲得與會者的回饋與建議，可做為後續技術發展與實務應用方向的參考。

於 OpenACC User Workshop 分享工作成果「GPU-Accelerated Particle Tracking on Unstructured Meshes Using OpenACC」之演講投影片

3.2 AI Scientist

在本次在本次 SCA/HPC Asia 2026 中，引起熱烈討論的一個主題是 Prof. Hiroaki Kitano 提出的「AI as a Scientist」。簡單來說，過去我們談的多是「AI for Science」，也就是把機器學習當作輔助工具；但現在越來越多團隊在發展所謂的 Agentic Science。這類系統結合大型語言模型（LLMs）與自主代理人，逐漸具備端到端完成整個研究

流程的能力：從提出假說、撰寫並優化實驗程式碼、分析模擬數據，到撰寫論文並通過同儕審查投稿。

在這個背景下，Prof. Kitano 提出的「Nobel Turing Challenge」更明確地設定了目標：發展高度自主的 AI 與機器人系統，使其能夠完成重大科學突破，甚至達到或超越諾貝爾獎等級的成就。基於這個趨勢，個人整理了近期該領域具有代表性的關鍵文獻，撰寫了一份簡要 Review Report（詳見附錄一）。內容從「AI Scientist」的宏觀願景出發，並進一步回顧其在純數位研究、HPC 系統自動化測試，以及延伸至實體實驗室等不同場域的最新進展。

3.3 Mixed-precision Computation

本工作坊深入探討 Ozaki-scheme 及其相關的數值分解策略。這項技術的核心在於透過誤差自由轉換（Error-Free Transformation, EFT），將原本必須在 FP64 高精度下執行的矩陣運算，拆解並分流到吞吐量更高的低精度單元完成。這對 FugakuNEXT 等下一代超級電腦特別重要，因為它證明即便硬體原生 FP64 單元有限，也能透過演算法模擬出與雙精度等效甚至更高精確度的結果。雖然這種方法會增加總體計算量（FLOPs），但由於低精度單元的運算效率極高，最終仍能縮短總執行時間並提升能效比。

在硬體應用方面，討論聚焦於 NVIDIA Tensor Cores 如何在科學模擬中發揮 AI 導向硬體的潛能。除了傳統 FP16，分析也涵蓋動態範圍更高的 BF16 以及新一代高效能 FP8 的實務應用。透過混合精度運算，開發者可以有效緩解 HPC 系統面臨的 Memory Wall 瓶頸。低精度數據佔用的位元數較少，能大幅降低記憶體頻寬與快取壓力，使程式從記憶體受限區轉向計算受限區，顯著提升大規模矩陣運算的吞吐量。

在演算法優化方面，討論包含 Iterative Refinement 與多精度動態切換策略。這種方法會在運算的關鍵步驟(如殘差計算)使用 FP64 高精度，其餘耗時計算則交由低精度處理，利用近似解快速收斂至精確解。在 CFD（計算流體力學）及複雜線性代數求解器中，結合 Perci-AD 或 Verificarlo 等自動化分析工具監控數值穩定性，可實現精確的動態精度分配。這種「軟體定義精度」策略不僅維持科學計算的物理正確性，也降低系統部署與維護門檻，在異質運算環境下達成速度與精度的平衡。

更多細節可參見附錄二（Mixed-Precision Computing）。

3.4 Fugaku-NEXT 與次世代 HPC-AI 支援中心（HAIRDESC）

日本已正式啟動國家旗艦級計畫 「Fugaku-NEXT」，目標是在 2030 年前後投入運行。該計畫由 RIKEN Center for Computational Science（R-CCS） 負責系統研發與總體設計，並由 Research Organization for Information Science and Technology（RIST） 負責資源分配與應用推廣。計畫核心在於將傳統的純 CPU 節點架構，轉型為 HPC 與 AI 深度融合的異質運算平台。

在供應鏈方面，富士通(Fujitsu）主導系統整合與次世代 CPU 設計，同時納入 NVIDIA 作為 GPU 技術合作夥伴，確保硬體具備 高頻寬記憶體 和強大的低精度張量運算能力，以支援 Exascale 級的大規模物理模擬及生成式 AI 運算需求。

為了克服從 CPU 主導轉向 GPU 主導的結構性挑戰，日本政府也發起了 「次世代 HPC-AI 支援計畫」，由 RIST 成立次世代 HPC-AI 研究開發支援中心（HAIRDESC）。該中心聯合筑波大學、東京大學及東京科學大學等學術單位，在 2025 至 2030 年間建立跨單位的技術支援體系。HAIRDESC 的核心任務不僅是硬體維護，更

著重於建立標準化 GPU 開發流程，協助國內研究機構將既有的 CPU-based 數值程式，透過 OpenACC、CUDA 或 HIP 等程式模型順利遷移到 GPU 架構。這不僅有助 Fugaku-NEXT 的硬體算力被充分利用，也解決研究單位在異質運算人才與程式移植上的缺口。

在運作模式上，HAIRDESC 採取「技術密集型專案資助」的策略，將重點從單純資金補助轉向實際技術人力投入，主要包括三個路徑：

- 委託研究與程式開發：由 HAIRDESC 派遣資深工程師直接參與重點研究項目的程式轉換，確保核心科學應用能順利過渡到多 GPU 環境。
- 提供 Testbeds：透過補助經費，在合作大學部署測試型 GPU 系統，供社群在旗艦機正式上線前進行大規模擴展性測試。
- 人才培育補貼：資助研究人員參加 RIST 舉辦的 GPU 移植工作坊，以及提供長期技術諮詢服務。

透過這種結構化的支援模式，確保國家級運算資源在部署初期就能維持高利用率，並具備良好的軟體跨平台運行能力。

3.5 Slurm + Slinky 的跨環境整合

Slurm 是成熟的排程系統，GPU 管理和異構運算支援已經很成熟，並非近期新技術。然而，Slurm 的角色已經從單一機房內的節點分配工具，逐漸轉向 跨環境算力調度的控制中樞。隨著 AI 訓練對資源彈性的需求增加，傳統 HPC 調度器必須具備可隨需求調整的雲端資源能力和系統互通能力，才能在本地基礎設施與雲端資源之間動態分配算力。

過去 HPC 系統與雲端容器平台（如 Kubernetes）整合雖已討論多年，但在實務上仍面臨資源分配不夠精細、高速網路管理複雜，以

及容器虛擬化層可能造成通訊延遲等挑戰。近年來，隨著雲端基礎架構的優化成熟，這些性能瓶頸已顯著改善。此時，由 SchedMD 主導的 Slinky 專案成為重要突破。Slinky 並非重新設計排程演算法，而是將 Slurm 容器化運行，並建立與 Kubernetes 資源的直接連接。這讓管理單位既能保留熟悉的 Slurm 指令操作，也能在雲端環境中隨需求動態擴展或釋放 GPU 節點，大幅降低 HPC 系統與雲端資源間的摩擦。

在產業層面，如果 NVIDIA 收購 SchedMD 的消息屬實，將代表 GPU 供應商開始進入排程層的垂直整合，未來效能優化不再只局限於硬體或驅動層，也會延伸到排程策略與資源配置邏輯。對超級電腦中心而言，核心關注可能已從技術新穎性轉向架構轉型對長期策略的影響。Slurm 與 Slinky 的組合象徵 HPC 調度器正演變為混合雲算力的統一控制中樞，將改變未來算力採購與基礎設施規劃，使算力調度更靈活、更具彈性。

3.6 AMD ROCm 架構轉型：TheRock 模組化建構體系

在技術定位上，AMD ROCm（Radeon Open Compute） 是 AMD 支撐 GPU 執行高效能運算（HPC）和 AI 工作負載的開放軟體平台。它由驅動程式、編譯器、執行時環境和各類運算庫組成，核心依賴 HIP（Heterogeneous-computing Interface for Portability），讓開發者能在 AMD 的伺服器與桌面 GPU 上直接使用 PyTorch、TensorFlow 等深度學習框架。ROCm 的長期價值在於，它提供了一套開源且高效的工具鏈，可作為 CUDA 生態之外的異質運算選項。

在近期技術上，AMD 在 ROCm 7.x 系列引入了 TheRock（The HIP Environment and ROCm Kit）。TheRock 並不改變 ROCm 的運算功能，而是重新整理了 ROCm 的軟體結構與安裝流程。過去 ROCm

的安裝與更新流程過於集中，套件也不夠靈活；TheRock 採用統一的 CMake 架構，將 HIP 與各類加速庫模組化組合，開發者可以自由選擇所需元件，減少多餘的軟體元件，同時提升在不同 Linux 發行版和 Windows 環境的兼容性。

TheRock 目前正在 ROCm 7.9 技術預覽中測試，預計在 2026 年中會取代原本的建構方式。它的技術優勢包括自動化建置、跨系統編譯，以及簡化下游專案（如 PyTorch）的整合流程。對高效能運算中心來說，這帶來兩大好處：第一，提高 ROCm 功能迭代與安全維護的效率；第二，支援範圍從大型伺服器 GPU 到桌面 GPU，甚至 CPU 的 AI 加速功能，讓算力資源調度與測試環境部署更靈活。

3.7 ADAC Applications & Benchmarks 工作小組

為了了解各國際計算中心在 Joint Benchmarks 的實務運作，並提升本中心在 ADAC 的實質參與度，本次會議中我代表中心參與了 Applications & Benchmarks 工作小組。針對不同數值模式，除了傳統的運算效能評比外，數值精度、能效比以及記憶體利用率等多維度指標，也都是小組在評估與優化核心應用程式時的重要依據。

在 GPU 計算方面，小組特別討論了 Eigenvalue 運算效率偏低，以及 FFT 在大規模擴展時受限等長期技術問題。同時，也並討論在 Fortran 與 OpenACC 等傳統科學開發環境下，ROCm 生態仍不夠成熟，對既有科學與工程等應用程式移植造成一定難度。此外，Mixed/Reduced Precision Computing 的實務應用也是各中心關注的焦點。針對上述議題，小組已建議在下次 ADAC 會議邀請相關領域專家進行專題演講，以利後續技術交流與對接。

在應用範疇方面，目前標竿測試仍以傳統科學與工程應用為主，例如 FLASH、GROMACS、GRONOR 等。不過，隨著 AI 快速發

展，如何將 LLM（大型語言模型）納入測試也成為熱門討論議題。雖然 AI 指標尚未完全標準化，小組建議可優先聚焦具差異化的測試方向，例如評估在本地語言或特定產業資料下進行模型微調時的訓練效率與資源成本，或透過新一代編譯框架，將傳統 HPC 計算與 AI 訓練流程整合在同一套工具鏈中，優化整體軟體生態。

透過本次參與，希望能協助本中心在 Applications & Benchmarks 小組中建立更實質且長期的投入。目前我正在彙整會議資料，並與中心的科學計算與 LLM 團隊討論，評估參與 Joint Benchmarks 所需的人力與時間成本，以規劃最合適的合作模式。期盼未來相關標竿測試數據也可作為設備採購與計畫提案的重要科學依據，同時為開發者提供關鍵的效能參考。

4.出國效益

本次差旅參加了 SCA/HPC Asia 以及 18th ADAC Symposium & Workshop，涵蓋高效能運算與加速運算的技術與應用發展。總體來說，透過 SCA/HPC Asia 的議程與技術交流，我們掌握了國際 HPC 的最新發展趨勢，包括 HPC 與 AI 的整合、異質運算架構的普及，以及效能與能效並行評估的做法。其中涉及的軟體生態建置、排程系統設計與應用程式優化方式，對本中心後續資源規劃與系統升級提供了直接參考。

在 ADAC 會議中，個人參與了 Applications & Benchmarks 工作小組，深入了解 Joint Benchmarks 的設計理念與運作流程，包括效能、數值精度與能效指標的評估方法，以及如何將測試結果作為系統規劃與評估依據。透過實際參與，也觀察到不同中心在應用程式選擇、性能分析與多維度評估的策略，這些經驗可作為本中心未來建立 Benchmark 評估機制、規劃算力資源與測試流程的重要參考。

在技術面觀察方面，本次差旅重點整理如下：

第一，AI Scientist 的發展模式

部分研究團隊已將大型語言模型整合到科學與工程研究流程中，用於文獻整理、參數搜尋與實驗設計輔助。這表示 AI 任務和傳統計算程式可以在同一台 HPC 系統上運行，因此未來在算力規劃上，需同時考量這兩類應用需求。

第二，混合精度（Mixed-precision）運算應用

混合精度技術已在 AI 訓練與大型線性代數求解中應用，透過不同數值精度配置與迭代修正方法，在控制數值誤差的前提下提升運算效率，這方法可以用來提升現有科學程式的運算效率。

第三，Slurm 與 Slinky 的資源整合架構

排程系統逐步與容器技術及雲端環境整合，Slurm 結合 Slinky 的模式顯示排程管理可以延伸至本地與雲端的跨環境資源調度，這提供了規劃混合雲和管理資源的實際參考。

第四，AMD ROCm 7.x 的模組化建構架構（TheRock）

ROCm 新版本改為模組化組件建構，使不同運算庫與工具可以按需求整合，並改善跨平台部署流程。此發展對評估不同 GPU 軟體和安裝方式提供了重要參考。

綜合來說，本次差旅除了掌握 HPC 與加速運算的最新技術，也透過參與 Applications & Benchmarks 工作小組，深入了解國際 HPC 中心在 Joint Benchmarks 的設計、效能分析與多維度評估方法。這些經驗有助於本中心未來建立自己的 Benchmark 評估機制、優化應用程式效能，並規劃算力資源配置，為 HPC 與 AI 運算提供清楚的策略與依據。同時，我們也將與中心同事討論參與 Joint Benchmarks 所需的人力與時間成本，以規劃最合適的合作模式，提升本中心在 ADAC 的實質參與度。

國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

附錄一、

AI Scientist

饒駿頌

高效能計算與雲端技術組

1. 摘要

近年來，人工智慧在科學研究中的角色正在發生轉變：它已經從專注單一任務的「輔助運算工具（AI for Science）」，快速進化成由大型語言模型（LLMs）驅動的「自主科學代理（Agentic Science）」。最新研究顯示，新一代 AI 科學家（AI Scientist）系統能夠從頭到尾完成科學工作流程，包括提出假說、設計實驗、撰寫程式碼，甚至進行論文審查。

本報告回顧這個領域的核心文獻，說明 AI 系統的架構如何演進、不同 AI 系統如何協同工作，以及如何在科學研究中遵循和應用基本物理原則。同時，也整理了將這類自主系統部署在高速運算（HPC）平台上可能面臨的挑戰。這項技術的發展，可能驅動科學研究正朝向高度自動化，也將迎來人類與 AI 共同合作的新模式。

2. 從諾貝爾圖靈挑戰到代理化科學（Scientific Agents）

科學發現的自動化一直是人工智慧研究的重要目標之一。Kitano (2021) 提出了「諾貝爾圖靈挑戰（Nobel Turing Challenge）」，希望開發出能自主進行頂級科學研究、並產生重大發現的 AI 系統。這個願景指出，未來的科學研究可能突破人類認知的限制，形成所謂的「Science of Science」。

隨著大型語言模型（LLM）技術的進步，學界已進入代理化科學的階段。Wei et al. (2025) 和 Ren et al. (2026) 的研究工作中指出，與傳統把 AI 當作黑盒預測工具不同，基於 LLM 的科學代理能夠整合

領域知識、使用外部工具（如編譯器和運算環境），並能自我檢查與修正錯誤，大幅提高研究流程的自動化和可重複性。

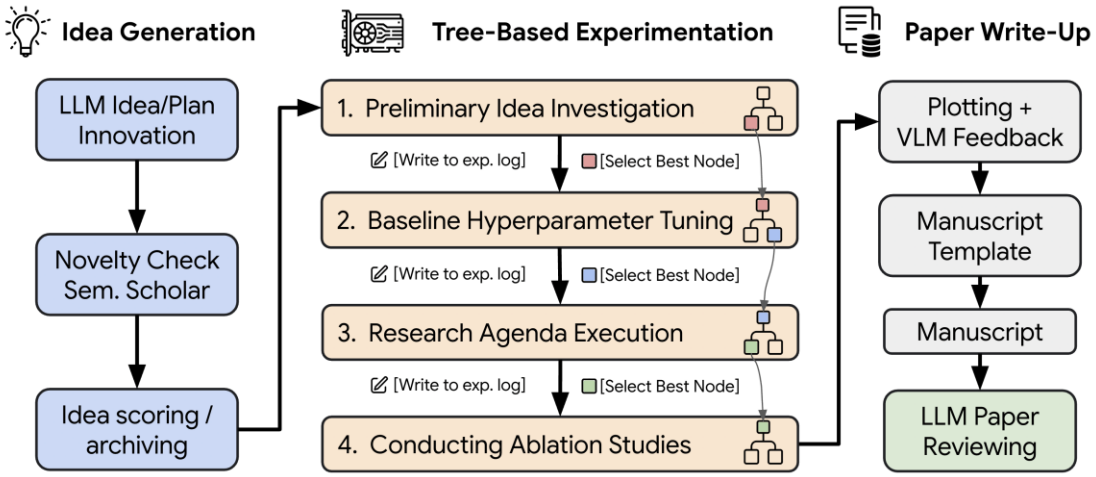

圖 1：The AI Scientist v2 的自動化科學研究流程。系統將科學研究分為四個階段：構想生成、以代理樹狀搜尋進行實驗探索、論文撰寫，以及 AI 自動審查。同時整合視覺語言模型形成視覺回饋機制，用於檢查並修正圖表與版面問題。（圖片來源：Yamada et al., 2025）

3. 核心系統架構與工作流程：以 AI Scientist 為例

為了實現高度自主的科學研究，研究團隊都提出了許多完整的自主研究系統。其中最具代表性的是 Sakana AI 團隊的 The AI Scientist 框架 (Lu et al., 2024; Yamada et al., 2025)。這個系統將科學研究拆解成一個高度自動化的循環流程：

1. **構想生成（Idea Generation）**：系統透過檢索和分析現有文獻，自主提出具新意的研究方向。
2. **實驗與實作（Experimentation）**：代理系統根據構想修改或生成程式碼，並在沙盒環境中執行運算和資料可視化。
3. **論文撰寫與審查（Paper Write-up & Review）**：系統將實驗結果生成標準學術手稿，再由內建 LLM 模組進行評

分和修正。

在第二代系統 The AI Scientist v2 中，研究團隊大幅重構了核心架構，引入了代理樹狀搜尋（Agentic Tree Search） 技術，並增加了專門的實驗管理員（Experiment Manager） 模組。在這套機制下，實驗管理員可以主動規劃多條不同的實驗路徑；如果某個程式碼分支在執行時遇到無法修復的錯誤，系統會自動回溯（Backtrack）並嘗試其他方案，而不會陷入死胡同。

這種探索與容錯能力讓系統不再依賴人類預先撰寫的程式碼模板，能夠在全新的研究領域中自主構建和修改底層程式碼，顯著提升了適應新研究領域的能力。此外，系統還整合了 視覺語言模型（VLM），形成視覺反饋迴圈。AI 能夠「看見」自己生成的數據圖表，自動檢查標籤重疊、圖例遮擋或排版問題，並依此修改繪圖程式，確保圖表內容和排版符合學術出版的標準 (Yamada et al., 2025)。

另外，Google 團隊提出的 AI Co-scientist 架構採用另一種不同的方式 (Gottweis et al., 2025)，基於科學方法設計了 「生成、辯論、演化（Generate, Debate, and Evolve）」 策略。具體來說，系統透過多個專責代理程式的協作來模擬真實的科研流程：

1. **生成代理（Generation Agent）**：檢索文獻並提出初步科學假說。
2. **排名代理（Ranking Agent）**：在多回合的模擬科學辯論中，讓不同假說兩兩競爭，逐步篩選出表現最好的假說。
3. **反思代理（Reflection Agent）**：像嚴格的同行評審一樣，檢查假說的正確性和可能的缺陷。
4. **演化代理（Evolution Agent）**：對篩選出的頂尖假說進行改良，融合其他想法、簡化概念或採用非常規思考方式進行迭代。

透過這種多代理互動的辯論與持續改進，系統能更嚴謹地檢驗科學假說的正確性和新穎性。

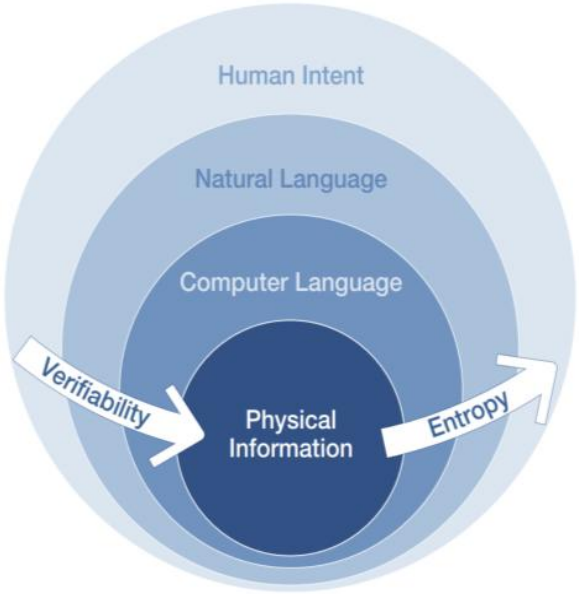

圖 2：自主科學代理系統的跨領域協調示意圖。在處理複雜科學計算時，AI 代理不只是語言模型，而是負責整體流程協調的核心角色。它整合人類科學家的領域知識、程式碼執行環境，以及底層的物理定律，讓研究過程在計算與推論上都符合物理原則。（圖片來源：Zhou et al., 2025）

4. 跨領域整合：協調語言、程式碼與物理法則

在除了純機器學習演算法的研究之外，將 AI 代理應用到更廣泛的自然科學與計算科學領域，已成為目前的重要發展方向。Zhou et al. (2025) 指出，自主代理在科學發現中的角色不只是工具，而是協調者，負責整合科學家知識、自然語言、程式碼與物理原理。

研究中也指出，傳統大型語言模型在處理複雜科學問題時，容易產生不符合物理現實的錯誤推論。新一代科學代理系統的做法，是將物理守恆定律與領域知識直接納入代理的決策流程與程式碼生成過程。也就是說，AI 在撰寫數值模擬或分析資料時，會受到數學模型與物理定律的限制，而不是單純依賴語言統計模式。具體而言，系統會為 AI 配備專屬的「物理驗證工具」（如呼叫外部數值求解器或符號運算引擎）並建立嚴格的反饋迴圈。當 AI 生成一段模擬程式碼

後，必須實際在物理引擎中編譯與執行測試；若運算結果違反了能量守恆或邊界條件，系統便會將錯誤日誌作為反饋傳回，強制 AI 代理重新修正其程式碼與數學推導，直到完全符合物理限制為止。這樣的設計有助於提升研究結果的可靠性。

這種整合方式對於處理高度複雜系統的領域特別重要，例如材料科學與流體力學等研究場景。

5. 高效能運算環境中的部署與挑戰

將具將上述能力的 AI 代理系統，從一般實驗伺服器部署到實際的高速運算叢集，是推動自動化科學計算的重要一步。Kotama et al. (2026) 針對將 The AI Scientist v2 應用於 HPC 環境，提出了具體的系統架構設計與挑戰分析。

文獻指出，HPC 環境的異質性與複雜度遠高於一般運算環境，因此 AI 代理必須具備高度的「環境透明度」，也就是能清楚理解並適應底層硬體與軟體條件。主要挑戰與對應作法包括：

- **異質架構與工具鏈適應**：代理系統必須能自動判斷目前使用的是哪一種類型的處理器（CPU 或 GPU 及其型號），選擇合適的編譯器編譯程式，並正確設定所需的數學運算與平行計算函式庫，讓程式能在該硬體環境下順利且有效率地執行。
- **本地端模型整合**：由於 HPC 環境常涉及未公開的研究資料與程式碼，研究者建議在本地部署大型語言模型，負責任務分派與程式碼生成，以降低資料外洩風險。
- **自動化效能測試與調校**：研究中也展示了 AI 代理在效能優化上的應用潛力，例如系統可以自行設計測試流程，評估 OpenBLAS 在不同數值精度組合下的效能表現，並分析在多核心環境中如何分配計算工作，找出較佳的執行配置。

整體而言，將 AI 代理導入 HPC 不只是擴充算力，而是要求系統能理解並管理複雜的運算環境，才能真正發揮自動化科學計算的效果。

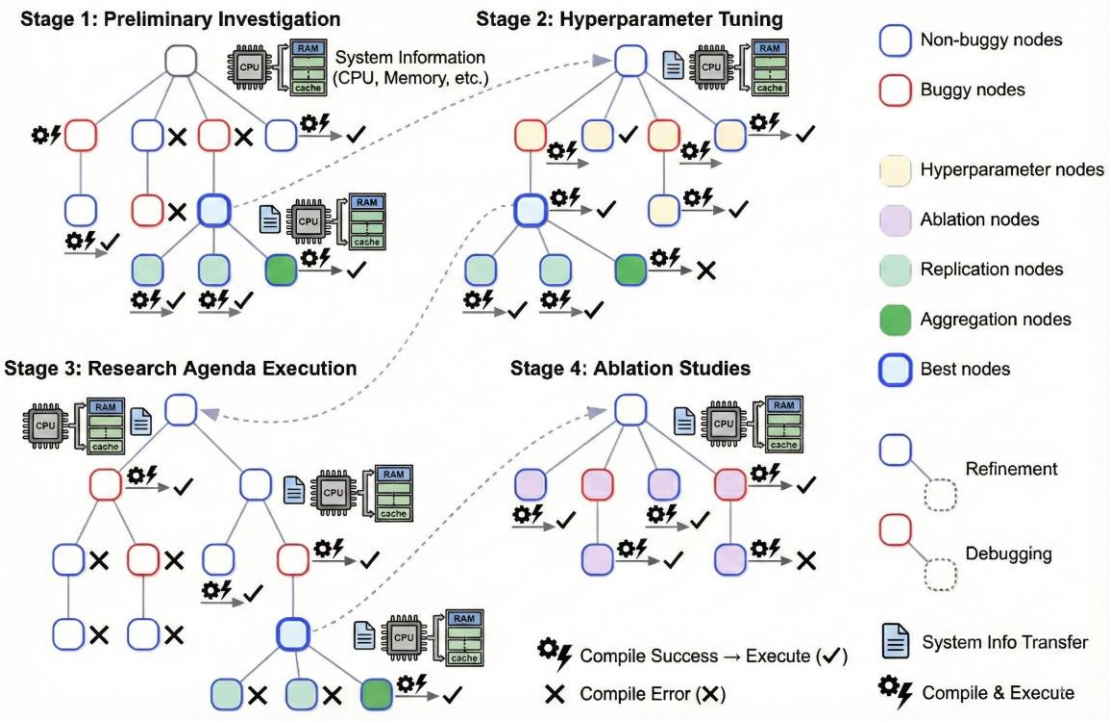

圖 3：HPC 擴展系統架構總覽。AI 代理於高效能運算環境中的自動化執行迴圈。系統透過彈性調用後端 LLM，並反覆執行「環境探測、程式碼生成、編譯執行與日誌解析」的迭代流程，以確保最終產出的程式碼能無縫適應異質的 HPC 叢集架構。(圖片來源：Kotama et al., 2026)

6. 結論

人工智慧在科學研究中的角色，正從單純的運算加速工具，轉變為能參與研究流程的協作系統。綜合現有文獻可以看到，無論是理論架構還是實際系統實作，高度自動化的科學工作流程已經初步成形。

未來的研究重點，將放在提升代理系統在跨領域情境下的適應能力，確保生成結果符合物理與數學原則，以及解決在異質 HPC 基礎

設施中部署與運行的技術問題。隨著代理架構與本地端大型語言模型技術逐步成熟，自主科學代理有機會成為科學計算環境中的常規工具，進一步提升研究效率。

參考文獻

[1] Gottweis, J., et al. (2025). Towards an AI co-scientist. *arXiv preprint* arXiv:2502.18864.

[2] Kitano, H. (2021). Nobel Turing Challenge: Creating the engine for scientific discovery. *npj Systems Biology and Applications*.

[3] Kotama, T., Yokota, R., Mukunoki, D., Hoshino, T., & Katagiri, T. (2026). Proposal of The AI Scientist v2 for high performance computing with local large language models. In *HPC Asia 2026*.

[4] Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). The AI Scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint* arXiv:2408.06292.

[5] Ren, S., et al. (2026). Towards scientific intelligence: A survey of LLM-based scientific agents. *arXiv preprint* arXiv:2503.24047.

[6] Wei, J., et al. (2025). From AI for science to agentic science: A survey on autonomous scientific discovery. *arXiv preprint* arXiv:2508.14111.

[7] Yamada, Y., et al. (2025). THE AI SCIENTIST-V2: Workshop-level automated scientific discovery via agentic tree search. *arXiv preprint* arXiv:2504.08066.

[8] Zhou, L., et al. (2025). Autonomous agents for scientific discovery: Orchestrating scientists, language, code, and physics. *arXiv preprint* arXiv:2510.09901.

國家實驗研究院
NATIONAL INSTITUTES OF APPLIED RESEARCH

附錄二、

Mixed-Precision Computing

饒駿頌

高效能計算與雲端技術組

1. 摘要

高效能運算（HPC）過去一直以雙精度浮點數（FP64）作為科學計算的主要數值標準。不過，隨著人工智慧相關的運算需求成為晶片架構設計的主導因素，現代 GPU 與資料中心系統已明顯將硬體資源配置在低精度算力（如 FP16、BF16、INT8）。這種設計方向與傳統科學軟體對高精度運算的需求之間出現落差，導致許多硬體的理論峰值算力難以在實務應用中被有效發揮（Kashi et al., 2025b）。

本文整理混合精度運算的主要理論基礎與演算法發展，說明其如何在記憶體牆（Memory Wall）問題（Kashi et al., 2025a）下改善資料移動成本、提升整體能效，並更有效利用以 AI 為導向的硬體架構。同時，也討論 Ozaki scheme（Mukunoki, 2025；Uchino et al., 2025；Ozaki et al., 2025）、迭代改進法（Haidar et al., 2020；Higham & Mary, 2022），以及自動化多精度優化工具（Chen et al., 2026；Zhou et al., 2025）的最新發展與應用方向。

2. 從絕對精度到資源最適化

在傳統 HPC 設計中，「數值精度」一直被視為科學計算正確性的基準，雙精度浮點數（FP64）長期是主要標準。然而，隨著硬體生態系統的改變，這個假設開始面臨結構性的挑戰。以 NVIDIA 為代表的現代加速器架構為例，為了應對 AI 訓練與推論對高速運算的需求，晶片設計重點已經轉向低精度運算。在最新的 GPU 中，晶片面積大部分用於處理半精度（FP16/BF16）或整數（INT8）運算的單元

（如 Tensor Cores），這些單元的理論峰值吞吐量可比 FP64 高出 8 倍、16 倍甚至更多（Kashi et al., 2025b；Ozaki et al., 2025）。

這種算力的非對稱分布意味著，在現代超級電腦或資料中心中，超過 80% 的理論總算力來自低精度硬體（Kashi et al., 2025b）。如果科學軟體仍僅使用 FP64，將造成硬體資源閒置和能源效率低下，因此研究者需要重新考慮如何充分利用高精度與低精度硬體資源。

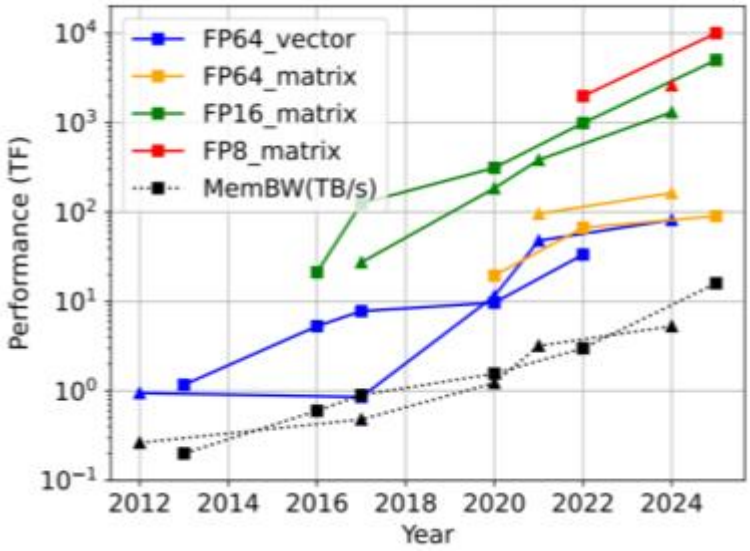

圖 1：GPU 硬體算力對比：FP64（高精度） vs. Tensor Cores（低精度）。觀察近年的 GPU 發展，傳統科學計算使用的 FP64 性能提升已趨於平緩，而為 AI 設計的低精度單元（Tensor Cores）則帶來了大部分算力增長。這表示，如果仍只使用 FP64 運算，超過 90% 的硬體算力可能無法被有效利用。（圖片來源：Kashi et al., 2026）

3. 混合精度的策略與系統效益

混合精度的核心想法是把運算任務按「重要性和計算量」來分配，而不是單純降低精度。計算密集、對小誤差不敏感的基礎算子（如大規模矩陣乘法）會分配到高吞吐量的低精度單元；而修正誤差和保

持計算穩定性則保留在 FP64 高精度路徑中（Higham & Mary, 2022；Le Gallo et al., 2018）。這種「粗算加精修」的方式讓系統在維持科學計算精度的同時，突破單一精度運算的效能限制。

3.1 記憶體瓶頸的緩解

在現代異質運算架構中，計算單元的性能增長速度遠高於記憶體頻寬，導致許多應用受限於數據移動速度，也就是所謂的「記憶體牆」。低精度格式（如 FP16 或 INT8）只需要 FP64 四分之一到八分之一的儲存空間，因此在相同頻寬下，硬體可以以更高速度將數據送入計算單元。這讓程式能更有效利用計算單元，不再被記憶體傳輸速度限制（Kashi et al., 2025a）。

3.2 能效比的提升

在 Exascale 運算中，功耗已成為系統擴展的主要限制。低精度運算單元因電路設計較簡單，單次運算消耗的電力顯著低於 FP64 單元。實證研究顯示，混合精度技術可以在不增加功耗的情況下，換取數倍的浮點運算產出。對於長時間運算的氣候模型或大規模流體模擬而言，每單位電力能完成更多運算，也能降低資料中心營運成本並支援綠色計算（Chen et al., 2026；Le Gallo et al., 2018）。

4. 演算法突破：Ozaki scheme 與迭代改進機制

混合精度運算的效益不僅來自硬體速度，也依賴演算法的設計。

4.1 Ozaki scheme 與誤差自由轉換（Error-Free Transformation）

Ozaki scheme 提供了一種將高精度運算轉到低精度硬體的方法，其核心技術是誤差自由轉換（EFT）。這項技術會將原始浮點數矩陣

拆成多個數值範圍較小的子矩陣，確保每個子矩陣的數值都落在低精度硬體（如 FP16 或 INT8）的表示範圍內。運算時，系統先在低精度單元上平行完成多次子矩陣乘法，最後再在高精度環境中加總補償。這個流程能保留運算中的所有數值精度，使低精度硬體也能模擬出 FP64 或更高精度的結果（Uchino et al., 2025）。

簡單來說，Ozaki 方案的運作可以概括為「先拆解、後運算、再重組」。如圖 2 上所示，原始高精度矩陣被切分成多個 Segments，確保每段數值不超過低精度單元的上限；接著，如圖 2 下所示，這些分段被送入矩陣運算單元（例如 Tensor Cores）進行平行計算。這個流程把精度問題轉化為可平行處理的計算量分配。

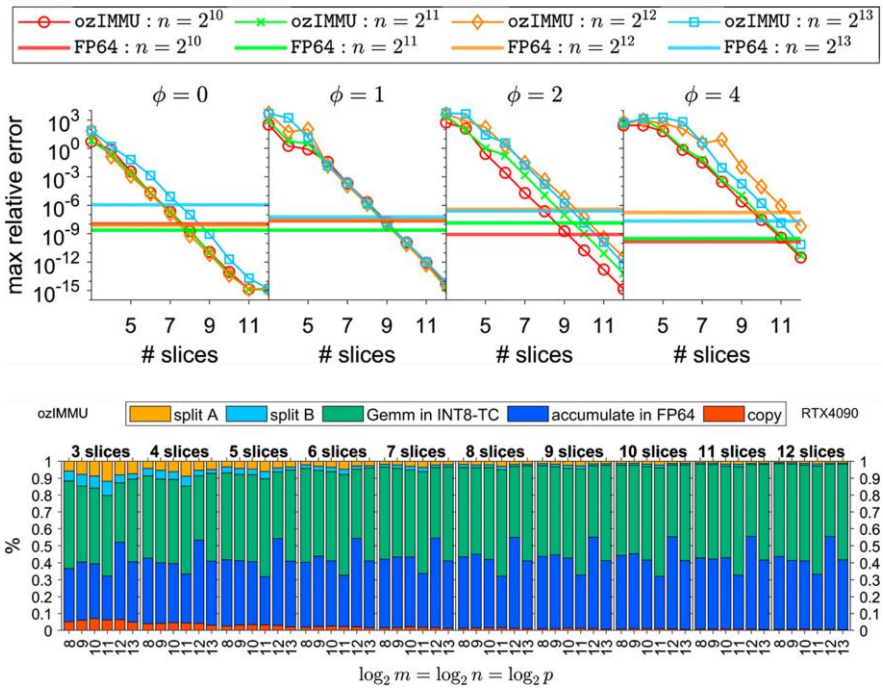

圖 2：Ozaki 方案矩陣分解與硬體運算流程。上圖顯示如何用 EFT 將高精度矩陣拆成多個低精度子矩陣；下圖展示這些子矩陣在 Tensor Cores 上的平行運算及最後的高精度加總。這種方法讓僅支援低精度的硬體，也能完成 FP64 或更高精度的計算結果（圖片來源：Uchino et al., 2025）。

最近的 Ozaki Scheme II 引入了 Chinese Remainder Theorem (CRT) 的整數運算，將浮點運算轉成多組獨立的整數同餘運算。這讓 NVIDIA Tensor Cores 在 INT8 模式下的效能更高，模擬 FP64 運算的速度比原生 FP64 硬體快約 7 到 10 倍（Ozaki et al., 2025）。

4.2 迭代改進法（Iterative Refinement）

迭代改進法用來解大規模線性方程組。方法是先用低精度算出初步結果，再用高精度修正誤差。演算法先在 FP16 等低精度下完成主要矩陣分解，然後用 FP64 計算誤差，多次迭代修正，最後得到和全 FP64 運算一樣精確的解。這種「低精度預測、高精度修正」的方法，可以在處理數萬階的大型矩陣時，把求解時間縮短到原先的三分之一到五分之一，同時保持數值穩定性，即使系統條件數較高，仍能可靠收斂（Haidar et al., 2020；Higham & Mary, 2022）。

5. 自動化精度優化與工具鏈

過去，混合精度部署對開發者來說門檻很高，需要判斷哪些變數可以降精度，擔心小幅精度損失會導致數值不穩定或結果錯誤。但新一代自動化工具正在將這個任務轉為可操作的工程流程。

5.1 Perci-AD 與自動微分技術

Perci-AD 框架利用自動微分（Automatic Differentiation）技術，自動追蹤程式中各運算變數對最終結果的影響（即「精度敏感度」）。透過計算梯度，系統能找出哪些模組對精度損失容忍度高，並自動生成最佳精度配置。開發者因此不需要手動測試每個變數，就能在保證數值穩定的前提下，更有效利用硬體運算能力（Zhou et al., 2025）。

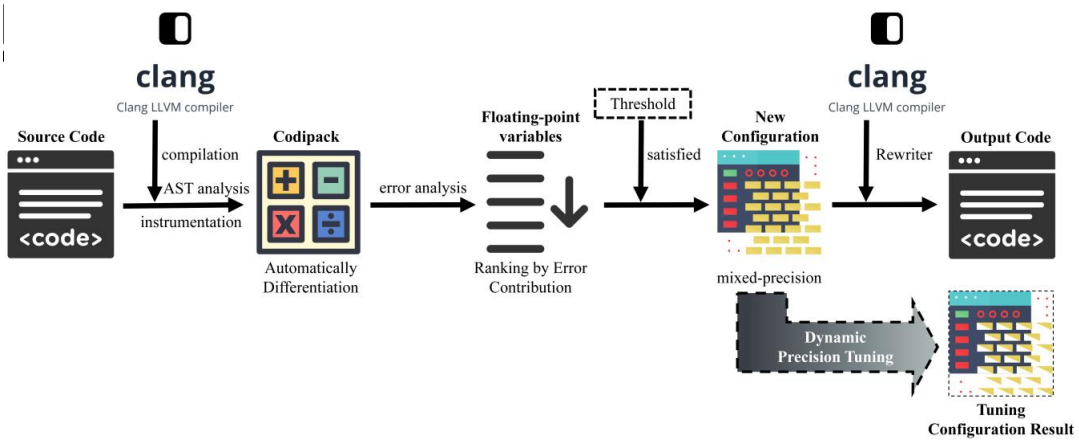

圖 3：Perci-AD 自動化多精度優化框架運作流程。系統透過自動微分技術追蹤程式中各運算變數對最終結果的影響程度，能自動找出對精度損失容忍度較高的模組，並自動配置合適的精度比例。這將過去需要專家手動測試的流程，轉為可執行的自動化工程流程。（圖片來源：Zhou et al., 2025）

5.2 Verificarlo 與複雜代碼的精度重構

對於像 CFD 或光譜元素法這類包含大量程式碼的系統，手動優化幾乎不可能。Verificarlo 提供 Dynamic Instrumentation 和隨機捨入分析功能，能模擬不同硬體精度下的數值行為，協助工程師快速找到會造成計算慢或累積誤差的程式部分」。透過這類工具的引導，例如 Nekbone 等科學應用已證明，在核心求解器中策略性地混用單精度與雙精度，不僅能縮短模擬時間，也簡化了代碼維護。這使得不具備深厚數值分析背景的工程師，也能可靠地進行精度優化，縮短從研究到生產部署的時間（Chen et al., 2026）。

6. 策略趨勢與發展路徑

觀察目前高性能計算與異質算力架構的演進，混合精度運算已經不只是學術議題，而是正在改變科學計算的基本方式。從前沿研究和大型系統實作中，可以歸納出幾個主要技術趨勢和發展方向：

6.1 精度敏感度分析的自動化

最新開發實務顯示，利用自動化工具來找出運算中可優化的部分已成為主流。例如 Perci-AD 框架透過自動微分技術，能判斷哪些程式模組對精度損失有較高容忍度（Zhou et al., 2025）。這種以數據為基礎的優化方式，正在取代過去依賴專家經驗的手動測試，降低了技術轉型的難度和風險。

6.2 開發流程與驗證標準化

軟體工程層面，像 Nekbone 或 Neko 這類應用已經開始建立標準化的精度驗證流程（Chen et al., 2026）。將精度測試納入 CI/CD 持續整合流程，可以在確保運算正確性的前提下，讓開發團隊更放心地下放精度設定。

6.3 演算法適配與硬體演進

隨著資料中心採用具備低精度加速能力的 AI 導向晶片（如 Tensor Cores），演算法研究也在調整方向，尋找方法釋放這些異質硬體的效能（Kashi et al., 2025b）。例如 Ozaki 方案提供了一種模擬高精度運算的方式，即使硬體原生只支援低精度，也能處理高精度任務（Mukunoki, 2025）。這種軟硬體協同優化的模式顯示，未來軟體競爭力將取決於對動態精度管理的適應能力。

7. 結論

混合精度運算代表了計算方式的轉變，其主要價值在於讓科學計算能與 AI 導向的硬體系統對接（Kashi et al., 2025b）。目前的技術成熟度和實證數據顯示，混合精度已具備生產環境部署的條件。靈活管

理精度不僅可以突破記憶體帶寬和能效限制，也將成為未來十年高性能計算與人工智慧結合的重要技術基礎。

參考文獻

[1] Chen, Y., de Oliveira Castro, P., Bientinesi, P., Jansson, N., & Iakymchuk, R. (2026). Enabling mixed-precision in spectral element codes. *Future Generation Computer Systems, 174*, 107990.

[2] Haidar, A., Bayraktar, H., Tomov, S., Dongarra, J., & Higham, N. J. (2020). Mixed-precision iterative refinement using tensor cores on GPUs to accelerate solution of linear systems. *Proceedings of the Royal Society A, 476*, 20200110.

[3] Higham, N. J., & Mary, T. (2022). Mixed precision algorithms in numerical linear algebra. *Acta Numerica, 31*, 347–414.

[4] Kashi, A., Koukpaizan, N., Lu, H., Matheson, M., Oral, S., & Wang, F. (2025a). Scaling the memory wall using mixed-precision – HPG-MxP on an exascale machine. In *SC '25: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (Article 281, pp. 1–17).

[5] Kashi, A., Lu, H., Brewer, W., Rogers, D., Matheson, M., Shankar, M., & Wang, F. (2025b). Mixed-precision numerics in scientific applications: Survey and perspectives. *arXiv preprint* arXiv:2412.19322v3.

[6] Le Gallo, M., Sebastian, A., Mathis, R., Manica, M., Giefers, H., Tuma, T., Bekas, C., Curioni, A., & Eleftheriou, E. (2018). Mixed-precision in-memory computing. *Nature Electronics, 1*, 246–253.

[7] Mukunoki, D. (2025). DGEMM without FP64 arithmetic — Using FP64 emulation and FP8 tensor cores with Ozaki scheme. *arXiv preprint* arXiv:2508.00441v3.

[8] Ozaki, K., Uchino, Y., & Imamura, T. (2025). Ozaki Scheme II: A GEMM-oriented emulation of floating-point matrix multiplication using an integer modular technique. *arXiv preprint* arXiv:2504.08009v3.

[9] Uchino, Y., Ozaki, K., & Imamura, T. (2025). Performance enhancement of the Ozaki Scheme on integer matrix multiplication unit. *The International Journal of High Performance Computing Applications, 39*(3), 462–476.

[10] Zhou, Y., Li, L., Liang, H., Zheng, C., Li, K., & Jiang, H. (2025). Perci-AD: Automated multi-precision optimization for high-performance computing. In *EDCS '25: Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area Education Digitalization and Computer Science International Conference* (pp. 852–859).

Domain-Adaptive Local Solver for Field Reconstruction in Proton Radiography

Chun-Sung Jao,[1] Chiung-Yin Chang,[2] Kun-Han Lee,[1] Hsiao-Hsuan Lin,[1] Han Hu,[1] Yen-Chen Chen,[1] Kentaro Sakai,[3] Yao-Li Liu,[4] Po-Yu Chang,[4] Hsiang-Yi Karen Yang,[2] Fang-An Kuo,[1] Tsung-Che Tsai,[1] and Yasuhiro Kuramitsu[5,6,7]

[1)]*National Center for High-Performance Computing, National Applied Research Laboratories, Hsinchu, Taiwan*
[2)]*Institute of Astronomy, National Tsing Hua University, Hsinchu, Taiwan*
[3)]*National Institute for Fusion Science, Gifu, Japan*
[4)]*Institute of Space and Plasma Sciences, National Cheng Kung University, Tainan, Taiwan*
[5)]*Graduate School of Engineering, Osaka University, Osaka, Japan*
[6)]*Institute of Laser Engineering, Osaka University, Osaka, Japan*
[7)]*Kansai Photon Science Institute, National Institutes for Quantum Science and Technology, Kyoto, Japan*

(*Electronic mail: C.-S. Jao - csjao@niar.org.tw, C.-Y. Chang - cychangastr@gapp.nthu.edu.tw)

(Dated: 25 February 2026)

Reconstructing multidimensional vector fields from path-integrated projection data is a fundamental challenge in high-energy-density physics, particularly when experimental sources exhibit spectral broadening and shot-to-shot jitter. We present a physics-informed deep-learning framework that addresses this ill-posed inverse problem by formulating global reconstruction as an aggregation of local inference tasks. By training a neural network on single-particle trajectories in randomized uniform magnetic fields, we develop a 'local solver' capable of zero-shot generalization to complex magnetohydrodynamic (MHD) topologies. To handle non-ideal laser-driven proton sources, we introduce a spectral out-of-distribution (OOD) filter that rejects non-physical inputs and prevents extrapolation errors. We demonstrate high reconstruction fidelity ($R > 0.99$, NRMSE $< 3\%$) across diverse spectral conditions. Furthermore, we provide a verifiable a priori metric for uncertainty quantification based on the in-distribution fraction of the source spectrum. This approach can be directly tuned to match experimental hardware, such as stacked nuclear track detectors, establishing a robust and parallelizable pipeline for quantitative plasma diagnostics.

I. INTRODUCTION

High-energy-density (HED) laboratory plasmas, generated when intense lasers interact with solid or gaseous targets,[1] exhibit rapidly evolving electric and magnetic fields that reproduce astrophysical phenomena such as shocks, magnetic reconnection, and various plasma instabilities.[2–7] Understanding the generation and evolution of these fields is a central goal of HED physics. Ion radiography (or proton imaging) is a powerful diagnostic for this purpose, mapping transient field structures by recording the deflection of high-energy protons traversing the plasma.[8–11] Upon reaching a detector, spatial variations in proton fluence encode integrated information about the electromagnetic fields encountered along the trajectories.

Despite its strong diagnostic capability, quantitative reconstruction of fields from proton radiographs remains a challenging ill-posed inverse problem. The problem is intrinsically nonlinear and under-constrained, because a radiograph contains only a two-dimensional projection of the cumulative Lorentz-force deflections accumulated along three-dimensional particle trajectories.[12,13] Different field configurations can produce indistinguishable fluence patterns, and the formation of caustics, where particle trajectories overlap, introduces singularities that are difficult for standard inversion techniques to handle.[12,14] Conventional approaches typically rely on regularized inversion or forward-fitting under simplified assumptions, such as monoenergetic proton sources.[15–17] However, practical laser-driven sources (e.g., TNSA) exhibit broadband energy spectra and substantial shot-to-shot energy jitter,[18,19] which violate these idealized assumptions and significantly reduce reconstruction accuracy.

In recent years, Scientific Machine Learning (SciML) has emerged as a transformative alternative for solving high-dimensional inverse problems where classical algorithms struggle.[20–24] By incorporating physical constraints into neural architectures, deep learning models can approximate complex non-linear mappings that are computationally expensive or analytically intractable. In the context of plasma diagnostics, this approach has enabled the automation of complex analyses, from optimizing laser-target parameters[25] to inferring plasma properties directly from experimental data.

Field reconstruction strategies generally follow two paradigms.[26–28] The first treats the radiograph as an image, employing Convolutional Neural Networks (CNNs) to map fluence patterns directly to field profiles.[26] While effective at capturing spatial correlations, these "image-to-image" models often struggle to generalize outside their training distribution (e.g., to new field topologies). The second strategy, the "inverse particle motion" approach, learns the local physics of particle deflection.[27,28] This framework acts as a general-purpose "local solver," making it inherently more robust to topological changes in the global field.

In this work, we extend the inverse particle-motion framework to address a key bottleneck for applying machine learning to experiments: the domain gap between idealized training data and non-ideal experimental conditions. In particular, we focus on magnetic field reconstruction in the presence of spectral broadening and shot-to-shot energy jitter. In contrast to previous studies that assume monoenergetic particle sources,[15,17,28] we introduce a spectral out-of-distribution (OOD) filtering mechanism. This approach enables a network trained only on simple, local particle–field interactions to robustly reconstruct complex global magnetic topologies, such as MHD vortices, even when the driving source is noisy and polychromatic. Furthermore, by explicitly quantifying the "valid data fraction" determined by the source spectrum, our framework provides an a priori metric for reconstruction reliability. This directly addresses the important requirement for uncertainty quantification in AI-driven diagnostic methods.

II. PHYSICS-INFORMED LEARNING FRAMEWORK

The proposed framework addresses the inverse problem by decomposing global magnetic field reconstruction into many local inference tasks. As shown in Fig. 1, the workflow consists of three main stages. First, synthetic proton radiographs are generated from two-dimensional magnetic field profiles obtained from MHD simulations, which serve as ground truth. Second, instead of training on full radiograph images, a neural network is trained on the elementary interaction between a single charged particle and a uniform local magnetic field. This training strategy produces a domain-generalizable local solver. Finally, the trained network is applied to synthetic radiographs to infer spatially resolved magnetic field distributions and recover the underlying global field structure.

To ensure robustness for realistic laser-driven proton sources with broadband spectra, a spectral OOD filter is incorporated during inference. This filter rejects inputs that fall outside the training energy envelope, preventing extrapolation and improving reconstruction fidelity.

A. Experimental Geometry and Interaction Model

The physical model is based on a standard laser-driven proton radiography setup (Fig. 1 (a)). A proton source at the origin $(0,0,0)$ emits particles that traverse a magnetic interaction volume centered at $z = 8$ mm with a transverse extent of 0.512×0.512 mm^2 and thickness $L_z = 0.256$ mm. Within this "local" interaction zone, the magnetic field is assumed to be z-invariant. A detector (CR-39) located at $z = 48$ mm records the final phase-space coordinates $\mathbf{y} = [X, Y, V_x, V_y, V_z]$ for each particle. The inverse problem seeks to map this detector vector $\mathbf{y}$ back to the local field parameters $\mathbf{x} = [B_x, B_y]$ and the particle's entry coordinates $[X_{\rm in}, Y_{\rm in}]$.

B. Zero-Shot Evaluation Benchmark (MHD Vortex)

To assess the model's ability to generalize to unseen topologies (zero-shot generalization), we utilize magnetohydrodynamic (MHD) Orszag–Tang vortex simulations[29] as a ground-truth benchmark (Fig. 1 (b)). Generated using the PLUTO code[30], these profiles evolve from simple initial conditions into highly complex turbulent structures containing shocks and current sheets.[31,32]

Data extracted at nondimensional times $T = 1, 2, 3$ are mapped to physical units ($\approx [-3, 3]$ T) on a 512×512 grid, as shown in the ground-truth profiles of Fig. 2 (a). Crucially, none of these topological structures are included in the training dataset, ensuring that successful reconstruction relies on learning the underlying physical laws rather than memorizing field patterns.

C. Spectral Noise and Source Modeling

Realistic laser-driven proton sources introduce inherent uncertainty into the inverse problem through spectral broadening and shot-to-shot energy jitter. We model the source energy spectrum $S(E)$ as a Gaussian distribution $\mathcal{N}(E_0, \Delta E^2)$. To systematically evaluate the robustness of our method, we construct a test suite consisting of 15 different source configurations:

- **Energy jitter (domain shift):** mean energy $E_0 \in 8, 9, 10, 11, 12$ MeV.
- **Spectral broadening (noise):** energy width $\Delta E \in 1, 2, 4$ MeV.

For each configuration, synthetic proton radiographs are generated by propagating 16×512^2 particle trajectories through magnetic fields imported from MHD simulations using a Boris pusher with time step $\Delta t = 10^{-13}$ s. The resulting synthetic radiographs, which incorporate spectral broadening and serve as the model's input, are illustrated in Fig. 2 (b). This dataset allows for a rigorous test of the model's domain generalization capabilities under non-ideal source conditions."

D. Neural Architecture and Domain Randomization

Instead of training on global images, we employ a multi-head perceptron (MLP) that is trained on single-particle trajectories.

a. Training Data via Domain Randomization: We generate a training dataset of 50,000 samples by propagating protons through randomized uniform magnetic fields, where B_x and B_y are sampled independently and linearly from a uniform distribution $\mathscr{U}[-5, 5]$ T. This domain-randomization strategy encourages the network to learn the fundamental Lorentz-force relationship $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$, instead of overfitting to particular geometric field structures. Crucially, the training set contains no spatial correlation or topological information; the network never 'sees' a vortex, a shock, or a gradient during the learning phase.

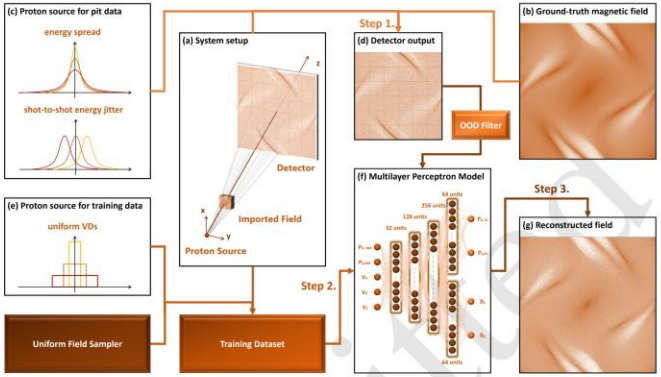

FIG. 1. Schematic of the physics-informed reconstruction pipeline. The framework bridges the domain gap between simple training data and complex experimental inference. (a–d) Synthetic validation data is generated using ground-truth MHD vortex simulations. (e–f) The neural network is trained as a generic "local solver" using only randomized uniform magnetic fields and single-particle trajectories. (g) During inference, the network reconstructs the global field topology from detector measurements by aggregating local predictions, optionally applying a spectral OOD filter to reject non-physical inputs.

b. Architecture: The network contains a shared feature extractor composed of three fully connected layers with 32, 128, and 256 neurons. The resulting latent representation is then divided into two task-specific heads, each with 64 neurons:

- **Field head:** predicts the local magnetic field vector (B_x, B_y).
- **Trajectory head:** predicts the particle entry coordinates $(X_{\text{in}}, Y_{\text{in}})$.

All input features are normalized: V_z is scaled to $[0,1]$, while the remaining phase-space coordinates $[X, Y, V_x, V_y]$ are scaled to $[-1,1]$.

c. Spectral Variants: To study the trade-off between precision and recall, we train three model variants that differ only in their training velocity bandwidth: NN10-4 (10 ± 2 MeV), NN10-2 (10 ± 1 MeV), and NN10-1 (10 ± 0.5 MeV). Each model is trained for 2000 epochs using a batch size of 25 and a learning rate of 10^{-3}. Training minimizes a combined mean-squared-error loss, $L_{\text{total}} = L_{\text{field}} + L_{\text{position}}$, and all models reach convergence with validation loss below 10^{-3}.

E. Spectral OOD Management

A primary challenge in deploying neural local solvers to experimental data is the presence of OOD inputs. In this framework, we define OOD samples as particles whose kinetic energies fall outside the specific spectral envelope used during the domain-randomization training phase. Because deep learning models typically exhibit poor extrapolation behavior, processing OOD particles leads to non-physical field predictions. We explicitly address this by implementing a spectral OOD filter that acts as a physical gate, ensuring the model only performs inference on data within its high-fidelity training domain.

F. Inference with Spectral OOD Filtering

Global field reconstruction is performed by aggregating the local predictions of the network. A key innovation of this framework is the spectral OOD Filter. In accordance with the OOD management strategy defined in Sec. II E, we strictly enforce a validity condition to prevent the aggregation of extrapolated predictions. We strictly enforce a validity condition:

$$\text{If } v_i \notin [v_{\min}, v_{\max}]_{\text{train}}, \quad \text{discard particle } i.$$

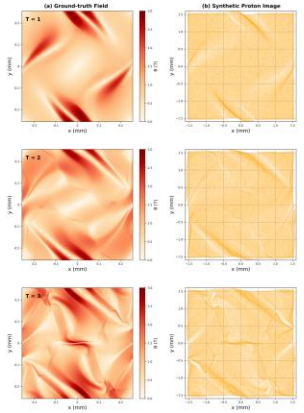

FIG. 2. Evaluation dataset representing complex topological structures unseen during training. (a) Ground-truth magnetic field profiles (B_x, B_y) extracted from MHD Orszag–Tang vortex simulations at time steps $T = 1, 2$, and 3. (b) Corresponding synthetic proton radiographs generated with a broad energy spectrum ($E_0 = 10$ MeV, $\Delta E = 4$ MeV), serving as the input for testing the model's domain generalization capabilities.

In practice, the corresponding energy acceptance windows are 8–12 MeV for NN10-4, 9–11 MeV for NN10-2, and 9.5–10.5 MeV for NN10-1.

For accepted particles, the predicted values are mapped onto the reconstruction grid using the predicted entry coordinates $(X_{\text{in}}, Y_{\text{in}})$. The final magnetic field value at each grid cell is obtained as the ensemble mean of all valid contributions assigned to that cell. In this way, the reconstruction becomes a selective regression problem, where the model gives priority to high-confidence (in-distribution) data rather than enforcing complete spatial coverage. While the ensemble mean provides a computationally efficient estimate for the final field values, more robust statistical approaches like the median or pixel-wise probability density functions could be employed in extreme noise environments. The OOD filter used in this work effectively removes the need for these complex estimators by cleaning the data before the aggregation step.

III. RESULTS AND DISCUSSION

A. Zero-Shot Generalization to MHD Topologies

Figure 3 demonstrates the model's ability to reconstruct complex, unseen magnetic topologies. Panel (a) displays the ground-truth B_x and B_y profiles from MHD simulations ($T = 1, 2, 3$), while Panel (b) shows the corresponding neural network reconstructions.

Importantly, although the network is trained only on single-particle interactions in simple uniform magnetic fields, it successfully recovers the multiscale structure of the Orszag–Tang vortex. This effective transfer from a randomized training domain to a structured physical domain demonstrates the domain-generalizable capability of the local-solver approach. The visual fidelity of the reconstructed filaments and shocks confirms that the model has captured the underlying particle-field dynamics rather than memorizing specific plasma morphologies, achieving true zero-shot generalization.

B. Ablation Study: Impact of OOD Filtering

The critical role of the spectral OOD filter during inference is demonstrated by comparing the first two columns of Fig. 3 (b). In the "No Filter" case, the NN10-4 model is applied to the entire polychromatic proton distribution. Without the gating mechanism, the model is forced to perform regression on particles with energies significantly outside its training range (10 ± 2 MeV).

As shown in the leftmost column of Fig. 3 (b), this leads to catastrophic extrapolation errors. Specifically, at $T = 3$, the reconstruction exhibits massive "white-out" artifacts where predicted values deviate wildly from the ground truth. Because the global field is reconstructed through ensemble averaging, even a small fraction of these out-of-distribution particles disproportionately distorts the mean field calculation. By contrast, the spectral OOD filter effectively masks these outliers, ensuring only physically valid, in-distribution predictions contribute to the final field map.

We quantify this improvement using the Pearson correlation coefficient (R) and the Normalized Root-Mean-Square Error (NRMSE), defined as:

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(B_i^{\text{pred}} - B_i^{\text{true}})^2}}{B_{\max}^{\text{true}} - B_{\min}^{\text{true}}}$$

, where N is the number of grid cells populated by valid particle trajectories. To ensure a physically meaningful comparison, the metrics are calculated exclusively on valid grid points, effectively masking cells with null values or those lacking in-distribution particle contributions.

Without the OOD filter, the reconstruction fidelity collapses: for $T = 1$ and 2, R drops to ≈ 0.3 with an NRMSE of $\approx 50\%$. By $T = 3$, the metrics reach a failure state with $R \approx 0.02$ and NRMSE $\approx 120\%$. Conversely, applying the spectral OOD filter restores high fidelity ($R > 0.99$, NRMSE $< 3\%$)

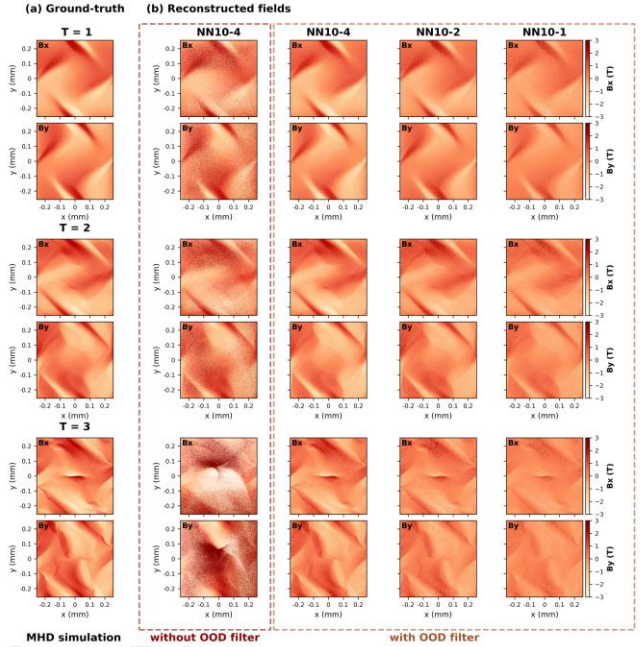

FIG. 3. Qualitative assessment of reconstruction robustness and the impact of OOD filtering. (a) Ground-truth MHD magnetic field profiles (B_x, B_y). (b) Reconstructed fields across different model configurations. The leftmost column (NN10-4, No Filter) shows significant artifacts due to spectral contamination. The subsequent columns (with OOD filter) demonstrate high-fidelity reconstruction, highlighting the trade-off between spatial coverage (Fill Factor) and the strictness of the spectral acceptance window (NN10-4 vs. NN10-1).

across all time steps, confirming the filter is a requirement for reliable inference in polychromatic environments.

C. Precision-Recall Trade-off

A comparison of the filtered models (NN10-4, NN10-2, NN10-1) reveals a fundamental trade-off between spectral acceptance (recall) and spatial coverage. While NN10-4 retains more particles to deliver a continuous field profile with superior coverage, the highly selective NN10-1 model produces

accurate but sparse maps (low fill factor).

Despite these differences in coverage, all filtered models maintain high accuracy across the 15 tested source configurations (Table I). Consistently maintaining $R > 0.99$ and NRMSE $< 3\%$ confirms that the local solver is robust to both source noise and spectral jitter, provided the OOD filter remains active. This point-wise agreement demonstrates that the system maintains quality regardless of the input source width.

TABLE I. Quantitative evaluation of model robustness against spectral non-idealities. Correlation coefficients (R) and NRMSE (%) are reported for the NN10-4, NN10-2, and NN10-1 architectures across 15 distinct source conditions (varying mean energy E_0 and spectral broadening ΔE). Metrics are calculated only on populated grid cells for the complex $T = 3$ MHD snapshot, demonstrating stability even under high-jitter conditions.

E_0 (MeV)	ΔE (MeV)	NN10-4 (R / NRMSE)	NN10-2 (R / NRMSE)	NN10-1 (R / NRMSE)
8	1	0.996 / 2.05	0.994 / 2.68	0.995 / 2.24
	2	0.996 / 1.93	0.994 / 2.66	0.995 / 2.13
	4	0.996 / 1.96	0.994 / 2.54	0.995 / 2.16
9	1	0.997 / 1.78	0.995 / 2.49	0.995 / 2.10
	2	0.997 / 1.81	0.994 / 2.46	0.995 / 2.09
	4	0.996 / 1.94	0.994 / 2.53	0.995 / 2.12
10	1	0.998 / 1.62	0.994 / 2.51	0.996 / 1.98
	2	0.997 / 1.75	0.995 / 2.43	0.995 / 2.08
	4	0.996 / 1.92	0.994 / 2.51	0.995 / 2.13
11	1	0.997 / 1.63	0.995 / 2.29	0.996 / 2.01
	2	0.996 / 1.76	0.995 / 2.42	0.995 / 2.07
	4	0.996 / 1.93	0.994 / 2.52	0.995 / 2.12
12	1	0.996 / 1.88	0.994 / 2.36	0.995 / 2.07
	2	0.996 / 1.83	0.994 / 2.45	0.995 / 2.10
	4	0.995 / 1.96	0.994 / 2.52	0.995 / 2.13

D. A Priori Uncertainty Quantification

To quantify the spatial completeness of the reconstruction, we define the Fill Factor (FF) as the ratio of grid cells containing at least one valid (in-distribution) particle prediction to the total number of cells. Figure 4 analyzes the relationship between the In-Distribution (ID) Fraction, the percentage of particles passing the OOD filter, and the resulting performance.

- **Fidelity is Independent of Coverage:** Panels (a) and (b) show that reconstruction accuracy remains stable regardless of the ID fraction, indicating that the OOD filter successfully decouples "data quantity" from "data quality."
- **Coverage is Predictable:** Panel (c) demonstrates that the FF scales with the ID fraction, allowing experimentalists to use the source spectrum as a proxy for expected spatial coverage.

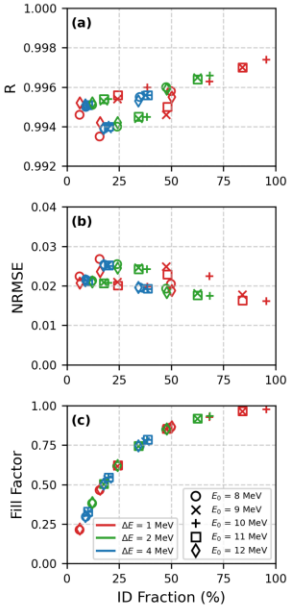

FIG. 4. Uncertainty quantification via the In-Distribution (ID) Fraction. The relationship between the particle acceptance rate (ID Fraction) and reconstruction quality. (a, b) Reconstruction fidelity (R, NRMSE) remains consistently high ($R > 0.99$) regardless of the ID fraction, confirming that the OOD filter successfully isolates valid data. (c) The spatial Fill Factor (FF) scales with the ID fraction, allowing for an *a priori* assessment of reconstruction coverage based solely on the source spectrum.

Because magnetic forces do not alter a particle's kinetic energy, the ID fraction is an intrinsic property of the source spectrum. This allows the metric to be calculated before performing inference, providing a practical "Go/No-Go" gauge. For instance, an ID fraction exceeding 25% consistently yields

a Fill Factor greater than 0.6, while a fraction above 50% ensures a Fill Factor exceeding 0.8 with reliable quality. This framework remains effective even with moderate electric fields, as high-velocity conditions in HED experiments ensure the detected energy closely represents the particle's energy during interaction.

IV. SUMMARY AND CONCLUSION

This study introduces a physics-informed domain-adaptation framework for solving the ill-posed inverse problem of electromagnetic field reconstruction. By shifting the learning paradigm from global image-to-image translation to a local solver approach, we address the critical challenge of reconstructing complex vector fields from path-integrated data corrupted by spectral noise.

Our results demonstrate that spectral OOD filtering is a key factor in achieving robust inference under realistic experimental conditions, together with the use of a uniform velocity distribution envelope during training. Standard neural networks, without these safeguards, suffer from large extrapolation errors arising from the domain gap between idealized training distributions and the polychromatic spectra of laser-driven sources. By integrating the OOD filter, we isolate the physically valid regime and reject uncertainty introduced by out-of-envelope inputs. The model consistently delivers high-fidelity reconstructions ($R > 0.99$, $NRMSE < 3\%$) with the best spatial coverage across all tested jitter conditions. While overly strict filters tend to reduce spatial coverage without improving precision, the NN10-4 model (trained on broader spectral data) provides an optimal balance between accuracy and coverage.

A major advantage of this framework is its capability for a priori uncertainty quantification. Since magnetic forces conserve kinetic energy, the In-Distribution (ID) fraction is determined solely by the source spectrum and can be calculated before performing the inference. From a practical standpoint, this spectral filter can be directly tuned to match the hardware configuration. For example, when using stacked nuclear track detectors like CR-39, the energy range of each layer is fixed by the stack design. Selecting an OOD filter window that matches the energy response of the detector layer ensures that the physical measurement and the neural network operate within the same domain. This provides experimentalists with a verifiable, ground-truth-free metric for assessing reconstruction reliability. Our results indicate that an ID fraction exceeding 25% consistently yields a spatial fill factor greater than 0.6, while a fraction above 50% ensures a fill factor over 0.8 with high-fidelity results.

Because the local solver operates independently of spatial gradients and global symmetries, the reconstruction framework is inherently independent of field topology. The Orszag–Tang vortex serves as a rigorous benchmark for multiscale structures, but the physical mapping learned by the network is fundamentally applicable to any arbitrary magnetohydrodynamic profile without further training.

From a methodological perspective, the local-solver architecture addresses two fundamental limitations of traditional global inversion methods. First, by treating single-particle dynamics independently, the network naturally resolves mathematical ambiguities arising from trajectory crossings and caustic formation. Second, the framework exhibits strong zero-shot generalization: although trained only on random uniform fields, it successfully reconstructs complex, multiscale magnetohydrodynamic vortex structures. Finally, because each local solver operates independently, the inference process is embarrassingly parallel. This enables high-throughput analysis on modern GPU clusters, establishing a robust and efficient machine-learning pipeline for high-energy-density physics diagnostics. While the current framework focuses on magnetic field reconstruction, future work will extend the 'local solver' architecture to explicitly disentangle concurrent electric and magnetic field contributions, leveraging the multi-head network's capacity to handle higher-dimensional phase-space mappings.

ACKNOWLEDGMENTS

This research is supported by the computing resources and infrastructure provided by the National Center for High-Performance Computing (NCHC), National Institutes of Applied Research (NIAR), Taiwan. This work is also supported by the NINS Program of Promoting Research by Networking among Institutions (Grant Numbers 01422301 and 01412302).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data and source code that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.18756432.

[1] M. Vogel, "Extreme physics: properties and behaviour of matter at extreme conditions, by jeff colvin and jon larsen," Contemporary Physics **55**, 256–256 (2014).
[2] Y. Kuramitsu, T. Moritaka, Y. Sakawa, T. Morita, T. Sano, M. Koenig, C. D. Gregory, N. Woolsey, K. Tomita, H. Takabe, Y. L. Liu, S. H. Chen, S. Matsukiyo, and M. Hoshino, "Magnetic reconnection driven by electron dynamics," Nature Communications **9**, 5109 (2018).
[3] T. G. White, M. T. Oliver, P. Mabey, M. Kühn-Kauffeldt, A. F. A. Bott, L. N. K. Döhl, A. R. Bell, R. Bingham, R. Clarke, J. Foster, G. Giacinti, P. Graham, R. Heathcote, M. Koenig, Y. Kuramitsu, D. Q. Lamb, J. Meinecke, T. Michel, F. Miniati, M. Notley, B. Reville, D. Ryu, S. Sarkar, Y. Sakawa, M. P. Selwood, J. Squire, R. H. H. Scott, P. Tzeferacos, N. Woolsey, A. A. Schekochihin, and G. Gregori, "Supersonic plasma turbulence in the laboratory," Nature Communications **10**, 1758 (2019).
[4] C.-S. Jao, S. Vafin, Y. Chen, M. Gross, M. Krasilnikov, G. Loisch, T. Mehrling, J. Niemiec, A. Oppelt, A. M. de la Ossa, J. Osterhoff, M. Pohl, and F. Stephan, "Preliminary study for the laboratory experiment of cosmic-rays driven magnetic field amplification," High Energy Density Physics **32**, 31–43 (2019).

[5] A. F. A. Bott, L. Chen, G. Boutoux, T. Caillaud, A. Duval, M. Koenig, B. Khiar, I. Lantuéjoul, L. Le-Deroff, B. Reville, R. Rosch, D. Ryu, C. Spindloe, B. Vauzour, B. Villette, A. A. Schekochihin, D. Q. Lamb, P. Tzeferacos, G. Gregori, and A. Casner, "Inefficient magnetic-field amplification in supersonic laser-plasma turbulence," Phys. Rev. Lett. **127**, 175002 (2021).
[6] Y. Kuramitsu, K. Sakai, and T. Moritaka, "Electron scale magnetic reconnections in laser produced plasmas," Reviews of Modern Plasma Physics **7**, 24 (2023).
[7] L. Gao, Y. Zhang, H. Ji, B. K. Russell, G. Pomraning, J. Griff-McMahon, S. Klein, C. Kuranz, and M. Wei, "Record magnetic field generation by short-pulse laser-driven capacitor-coil targets," Applied Physics Letters **127**, 094101 (2025).
[8] M. Schollmeier, S. Becker, M. Geißel, K. A. Flippo, A. Blažević, S. A. Gaillard, D. C. Gautier, F. Grüner, K. Harres, M. Kimmel, F. Nürnberg, P. Rambo, U. Schramm, J. Schreiber, J. Schütrumpf, J. Schwarz, N. A. Tahir, B. Atherton, D. Habs, B. M. Hegelich, and M. Roth, "Controlled transport and focusing of laser-accelerated protons with miniature magnetic devices," Phys. Rev. Lett. **101**, 055004 (2008).
[9] T. M. Ostermayr, C. Kreuzer, F. S. Englbrecht, J. Gebhard, J. Hartmann, A. Huebl, D. Haffa, P. Hilz, K. Parodi, J. Wenz, M. E. Donovan, G. Dyer, E. Gaul, J. Gordon, M. Martinez, E. Mccary, M. Spinks, G. Tiwari, B. M. Hegelich, and J. Schreiber, "Laser-driven x-ray and proton micro-source and application to simultaneous single-shot bi-modal radiographic imaging," Nature Communications **11**, 6174 (2020).
[10] D. B. Schaeffer, A. F. A. Bott, M. Borghesi, K. A. Flippo, W. Fox, J. Fuchs, C. Li, F. H. Séguin, H.-S. Park, P. Tzeferacos, and L. Willingale, "Proton imaging of high-energy-density laboratory plasmas," Rev. Mod. Phys. **95**, 045007 (2023).
[11] W. Fox, G. Fiksel, D. B. Schaeffer, and J. Griff-McMahon, "Proton deflectometry analysis in magnetized plasmas: Magnetic field reconstruction in one dimension," Phys. Rev. E **110**, 015206 (2024).
[12] N. L. Kugland, D. D. Ryutov, C. Plechaty, J. S. Ross, and H.-S. Park, "Invited article: Relation between electric and magnetic field structures and their proton-beam images," Review of Scientific Instruments **83**, 101301 (2012).
[13] C. Graziani, P. Tzeferacos, D. Q. Lamb, and C. Li, "Inferring morphology and strength of magnetic fields from proton radiographs," Review of Scientific Instruments **88**, 123507 (2017).
[14] J. M. Levesque and L. J. Beesley, "Reconstructing magnetic deflections from sets of proton images using differential evolution," Review of Scientific Instruments **92**, 093505 (2021).
[15] C. A. J. Palmer, P. T. Campbell, Y. Ma, L. Antonelli, A. F. A. Bott, G. Gregori, J. Halliday, Y. Katzir, P. Kordell, K. Krushelnick, S. V. Lebedev, E. Montgomery, M. Notley, D. C. Carroll, C. P. Ridgers, A. A. Schekochihin, M. J. V. Streeter, A. G. R. Thomas, E. R. Tubman, N. Woolsey, and L. Willingale, "Field reconstruction from proton radiography of intense laser driven magnetic reconnection," Physics of Plasmas **26**, 083109 (2019).
[16] C. Arran, C. P. Ridgers, and N. C. Woolsey, "Proton radiography in background magnetic fields," Matter and Radiation at Extremes **6**, 046904 (2021).
[17] J. Griff-McMahon, V. Valenzuela-Villaseca, S. Malko, G. Fiksel, M. J. Rosenberg, D. B. Schaeffer, and W. Fox, "Proton radiography inversions with source extraction and comparison to mesh methods," Phys. Rev. E **110**, 055202 (2024).
[18] M. F. Kasim, A. F. A. Bott, P. Tzeferacos, D. Q. Lamb, G. Gregori, and S. M. Vinko, "Retrieving fields from proton radiography without source profiles," Phys. Rev. E **100**, 033208 (2019).
[19] J. I. Apinaniz, S. Malko, R. Fedosejevs, W. Cayzac, X. Vaisseau, D. de Luis, G. Gatti, C. McGuffey, M. Bailly-Grandvaux, K. Bhutwala, V. Ospina-Bohorquez, J. Balboa, J. J. Santos, D. Batani, F. Beg, L. Roso, J. A. Perez-Hernandez, and L. Volpe, "A quasi-monoenergetic short time duration compact proton source for probing high energy density states of matter," Scientific Reports **11**, 6881 (2021).
[20] N. F. Y. Chen, M. F. Kasim, L. Ceurvorst, N. Ratan, J. Sadler, M. C. Levy, R. Trines, R. Bingham, and P. Norreys, "Machine learning applied to proton radiography of high-energy-density plasmas," Phys. Rev. E **95**, 043305 (2017).
[21] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," Acta Numerica **28**, 1–174 (2019).
[22] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics **378**, 686–707 (2019).
[23] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," Annual Review of Fluid Mechanics **52**, 477–508 (2020).
[24] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics **3**, 422–440 (2021).
[25] Y. Kuramitsu, T. Taguchi, F. Nikaido, T. Minami, T. Hihara, S. Suzuki, K. Oda, K. Kuramoto, T. Yasui, Y. Abe, K. Ibano, H. Takabe, C. M. Chu, K. T. Wu, W. Y. Woon, S. H. Chen, C. S. Jao, Y. C. Chen, Y. L. Liu, A. Morace, A. Yogo, Y. Arikawa, H. Kohri, A. Tokiyasu, S. Kodaira, T. Kusumoto, M. Kanasaki, T. Asai, Y. Fukuda, K. Kondo, H. Kiriyama, T. Hayakawa, S. J. Tanaka, S. Isayama, N. Watamura, H. Suzuki, H. S. Kumar, N. Ohnishi, T. Pikuz, E. Filippov, K. Sakai, R. Yasuhara, M. Nakata, R. Ishikawa, T. Hoshi, A. Mizuta, N. Bolouki, N. Saura, S. Benkadda, M. Koenig, and S. Hamaguchi, "Optimization of laser-driven quantum beam generation and the applications with artificial intelligence," Physics of Plasmas **31**, 053108 (2024).
[26] S. Kurochi, A. Mizuta, K. Sakai, K. Nikaido, C.-S. Tamaki, N. Jao, Y.-C. Chen, N. Ohnishi, and K. Y., "Numerical investigation on reconstruction of turbulent electromagnetic fields in laser-driven plasma using convolutional neural networks," Physics of Plasmas (under review).
[27] C.-S. Jao, Y.-C. Chen, F. Nikaido, Y.-L. Liu, K. Sakai, T. Minami, S. Isayama, Y. Abe, and Y. Kuramitsu, "Advancements in neural network techniques for electric and magnetic field reconstruction: Application to ion radiography," AIP Advances **14**, 025037 (2024).
[28] A. Mizuta, C.-S. Jao, S. Kurochi, F. Nikaido, K. Sakai, N. Ohnishi, Y.-C. Chen, Y.-L. Liu, N. Watamura, T. Minami, H. Suzuki, T. Hoshi, Y. Terachi, A. Nakano, and Y. Kuramitsu, "Numerical analysis of the evolution of hydrodynamic instabilities and vortex generation associated with collisionless shocks in laser-produced plasmas," Contributions to Plasma Physics **65**, e70019 (2025).
[29] S. A. Orszag and C.-M. Tang, "Small-scale structure of two-dimensional magnetohydrodynamic turbulence," Journal of Fluid Mechanics **90**, 129–143 (1979).
[30] A. Mignone, G. Bodo, S. Massaglia, T. Matsakos, O. Tesileanu, C. Zanni, and A. Ferrari, "Pluto: A numerical code for computational astrophysics," The Astrophysical Journal Supplement Series **170**, 228 (2007).
[31] P. Londrillo and L. Del Zanna, "High-order upwind schemes for multidimensional magnetohydrodynamics," The Astrophysical Journal **530**, 508 (2000).
[32] T. Tsai, H.-S. Yu, M.-S. Hsieh, S. Lai, and Y.-H. Yang, "Implicit predictor–corrector central finite difference scheme for the equations of magneto-hydrodynamic simulations," Computer Physics Communications **196**, 1–12 (2015).