ASC25 Student Supercomputer Challenge

Final Competition Notification

Dear ASC25 finalists.

The ASC25 Committee is pleased to congratulate all the qualifying teams of the ASC25 Student Supercomputer Challenge finals, which will take place onsite at Qinghai University.
Each team must comprise one adviser and five undergraduate students, maintaining the same lineup as originally registered during the Preliminary Round. If any changes are necessary, teams must submit a formal request, including a clear justification, sent via email to info@asc-events.org. We commend your achievement and look forward to your participation in the finals.

- During the finals, each team is required to design and set up their platform onsite while adhering to a **4 KW power consumption limit** and executing the approved application tasks outlined in this announcement. The cluster must consist of **at least three compute nodes, with a per-node power limit of 2 KW.**
- Each team must complete the Hardware Platform Equipment Form for the Final Stage (Appendix I) and submit it via email to TechSupport@asc-events.org by March 31, 2025. The ASC25 Committee will allocate the necessary resources to ensure the required equipment is available for the finals.
- The onsite finals will commence daily at 8:00 AM (UTC/GMT+8:00).
- **Additionally, unlike previous editions, this year's Final Competition will incorporate the Group Competition as a component of the overall score.** This new format aims to foster direct collaboration and knowledge exchange among teams, particularly in areas such as equipment assistance and application optimization.
- Due to adjustments to the Group Competition rules and increased task complexity in the current cycle, the competition format of Mystery application has been suspended for the ASC25 finals. The application is released within this notification.
- For detailed rules regarding the finals, please refer to Appendix II.

Schedule of the Finals

Date	Time	Content	
May 9	Check- in		
May 10	08:00-20:00	Announcement of contest rules, cluster building and testing	There is no limit on the runtime power consumption of the entire cluster during these periods.
May 11	08:00-20:00	Cluster building and testing	
May 12	08:00-18:00	Performance testing of HPL, HPCG, group competition and AlphaFold3 Inference	The runtime power consumption of the entire cluster should

May 13	08:00-18:00	RNA m^5C, Geant4, DeepSeek Inference	be **less than 4 KW**. The system platform cannot be rebooted or changed during May 12-13.
May 14	08:00-12:10	Team Presentation	
	15:00-18:30	The Awards Ceremony	
May 15	Check out		

* Note: Check-in must be no earlier than May 8, 2025, and check-out must be no later than May 16, 2025. If an extended stay is required due to long-distance travel or other reasons, please submit a request via email to info@asc-events.org.

Appendix I

ASC25 Student Supercomputer Challenge

Hardware Platform and Equipment for the Final Stage

Power Consumption Restrictions and Hardware Platform Description.

a) The primary objective is to design and optimize computing systems to achieve the best possible performance while running the approved applications within the 4 KW power consumption limit. Failure to comply with these power limitations will result in task invalidation for the affected team.
b) Each team must design their cluster based on the servers and components listed below. The ASC25 Committee will provide all required hardware except for GPUs. Teams may choose to use additional components at their own expense, except for the servers, which will be those provided by the ASC25 Committee. During the final stage (May 12–13, 2025), the system platform must remain unchanged and cannot be rebooted or modified. Each team is required to complete the Hardware Platform Equipment Form for the Final Stage (see table below) and submit it via email to techsupport@asc-events.org by March 31, 2025. Please note that the hardware configuration may be subject to minor adjustments due to unforeseen circumstances.
c) Competition equipment must remain powered on. Reboots only allowed for hardware failures with prior ASC25 staff notification. Hibernation/suspension modes prohibited, as standby states violate operational requirements.

Item	Name	Configuration	Note
Server	Dual Processor Server	CPU: Intel® Xeon® 6760P Processor * 2 Memory: 32GB * 16, DDR5, 6400 MT/s Hard disk: 480GB SSD SATA * 1	These items will be provided by the ASC25 Committee.
HCA card	NDR200	InfiniBand NVIDIA ConnectX®-7 NDR200	

Switch	GbE switch	10/100/1000 MB/s，24 ports Ethernet switch	
	NDR-IB switch	NVIDIA Quantum (TM)-2 NDR InfiniBand Switch, 64-ports NDR, 32 OSFP ports, unmanaged, P2C airflow (forward)	
Cable	Gigabit CAT6 cables	CAT6 copper cable, blue, 3 m	
	InfiniBand cable	InfiniBand NDR copper cable, OSFP port, compatible with the InfiniBand switch in use.	
GPU	The ASC25 Committee will NOT provide GPUs. Teams may bring their own GPUs for installation in the servers provided by the ASC25 Committee, with each server supporting up to 2 GPUs.		

Appendix II

ASC25 Student Supercomputer Challenge:

Technical Regulation and Evaluation Criteria for the Final Stage

Rules of the final stage:

1. The use of optimization methods specific to certain parameters or input data sets is strictly forbidden.
2. If any changes are made to the algorithm, the revised version must maintain mathematical equivalence to the original.
3. Violation of any rule mentioned above will result in a zero score being assigned for the corresponding task.
4. Carefully constructing the cluster is essential. Any damage to the server may incur a penalty of up to 20 points for the team, as determined by the ASC25 Committee.

Note:

Teams may consult the ASC25 Committee in advance regarding any uncertainties about whether a specific optimization method complies with the competition rules. The ASC25 Evaluation Committee will review such inquiries and provide a decision prior to the commencement of the final competition. Once the competition begins, no further explanations will be provided if an optimization method is deemed ineligible by the ASC25 Evaluation Committee.

Awards and prizes

Finals		
Award Name	**Bonus (CNY)**	**Rules**
Champion	100,000	The team that achieves the highest total score will be considered as the winner. **Important: The Group Competition results will be included in the final score calculation.**
Silver	50,000	The team that ranks second in the total score. **Important : The Group Competition results will be included in the final score calculation.**
e Prize	27,182	The team that achieves the highest score in AlphaFold3 Inference will be considered the winner of this prize.
The Highest Linpack	10,000	The team that achieves the highest score in HPL benchmark.
Application Innovation	3*10,000	The team that ranks first in the score in RNA m^5C, Geant4, and DeepSeek Inference, respectively. Note: (1) If a team wins any of the Champion, Silver, e Prize, or the highest Linpack awards, they will not be eligible

		for the Application Innovation Award. In such cases, the award will be granted to the team that ranks directly below them, and so on. (2) If a team secures first place in two or more applications, they will share the prizes with the team that ranks directly below them in those applications.
Most Popular Team	2*5,000	The team receiving the highest number of votes through online and onsite voting will be recognized: • Teams from mainland China will have their votes counted from WeChat voting and onsite voting. • Teams from outside mainland China will have their votes counted from X (Twitter) voting and onsite voting.
Group Competition		
Award Name	**Bonus (CNY)**	**Rules**
Group Competition Award	20,000	1. **Group Formation**: The groups will be announced prior to the finals, with each group consisting of five teams and required to include at least one team from outside mainland China. Any changes to the number of teams and other factors will be subject to the notification of the ASC25 committee. 2. **Assignment of Teams**: Each team will be assigned a unique ID number, generated randomly. Teams sharing the same ID number will be placed in the same group. 3. **Competition Application and Workloads**: The Group Competition application will be released within five days following the group formation. The workloads for the application will be announced on the first day of the competition. Teams within a group may collaborate to complete the application; however, each team must individually execute and complete the workloads on its own cluster. Clusters must not be operated directly or remotely by members of other teams in the group. The overall performance of each team within a group will contribute to determining the group's final result.

		4. **Collaboration and Performance Optimization**: Teams within a group are allowed to collaborate on hardware equipment, application compilation, debugging, optimization, and discussions. The output of each workload must undergo correctness verification to ensure validity. The primary objective is to achieve the shortest runtime for all workloads. 5. **Power Consumption Limitations**: Runtime power consumption must not exceed 4 KW. Failure to comply with this limit will result in task invalidation for the affected team. 6. **Results and Awards**: The results of the Group Competition will be announced on the morning of the second day of the finals. The winning group will be awarded the Group Competition Prize and corresponding bonuses. Bonuses will be distributed equally among all teams within the winning group. 7. **Impact on Overall Awards**: The results of the Group Competition will contribute to determining the Champion and Silver Award recipients.

Performance Optimization (90 points)

I. HPL performance optimization (6.5 points):

1. **Platform requirement:** Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **Goal:** Obtain the correct results while achieving the highest performance.
3. **Note:** All teams must commence the HPL benchmark immediately at the start of May 12. And HPL results must be submitted **before 11:00 AM**. Teams will only be able to get the application tasks of May 12 either after successfully submitting HPL results or after 11:00 AM.
4. **Software download:** http://www.netlib.org/benchmark/hpl/

II. Performance optimization of HPCG (6.5 points):

1. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **About run time:** The runtime of HPCG (version 3.0) must be a minimum of 1800 seconds (30 minutes), as reported in the output file. The Quick Path option is not permitted.
3. **Software download:** https://github.com/hpcg-benchmark/hpcg

III. Performance optimization of AlphaFold3 Inference (18 points):

1. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **Goal:** The AlphaFold3 Inference Challenge shares a similar objective with the Preliminary Round. However, for the Final Round**, only CPU-based inference is permitted**. During the competition, the ASC25 Committee will announce several new inference sequences onsite. Additionally, participants must strictly adhere to the following constraints:

   a) Modifying model weights like quantization or pruning and using precision below 16-bit is strictly prohibited. b) Adjustments to method parameters, such as recycling number, diffusion step, and others, are not allowed.
3. **Software download:** https://github.com/google-deepmind/alphafold3

IV. Performance optimization of RNA m^5C (18 points):

1. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **Goal:** The RNA m^5C site detection workload requires seamless integration of multiple computational tools and strategic parallelization of computing resources. During the ASC25 finals, the competition committee will deploy large-scale sample datasets, and teams must process all assigned datasets through the entire workflow. To ensure validity, all pipeline stages must execute without errors and produce accurate results. Correctness checks will be conducted, including assessments of precision metrics, biological correlation, and m^5C site counts. Teams must strive to minimize runtime while maintaining data accuracy and workflow efficiency. Any violation of academic integrity will result in immediate disqualification.
3. **Workload pipeline reference:** https://github.com/y9c/m5C-UBSseq

V. Performance optimization of Geant4 (18 points):

1. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **Goal:** Geant4 is a Monte Carlo simulation toolkit widely used in high-energy physics, space science, and medical applications. During the ASC25 finals, the competition committee will announce the workload for this challenge, which will be based on a modified ExampleB1 benchmark with various parameters using Geant4-v11.3.0. The primary goal is to minimize the runtime of the workload while ensuring that all submitted results pass correctness verification. Modifying any code related to method parameters is strictly prohibited. All parameters in the input files must remain unchanged, except for those specifically related to parallelization. Teams are allowed to modify the source code of the workload to optimize performance, but the source code of Geant4-v11.3.0 must remain unchanged.
3. **Software download:** https://www.geant4.org/download/11.3.0.html

VI. Performance optimization of DeepSeek Inference (18 points):

1. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification.
2. **Goal:** The goal of the DeepSeek Inference Challenge is to design and deploy an LLM (large language model) inference serving system using the **DeepSeek R1-Distill 32B**

model to process queries provided by the ASC25 committee on-site. The primary objective is to minimize inference time across various scenarios, focusing on both individual query latency and overall query throughput, while ensuring that all submitted results pass correctness verification. The serving system should leverage the OpenAI-compatible API to receive queries and efficiently handle concurrent query processing. Additionally, **the inference serving system is strictly limited to using CPUs for inference**. Participants must also strictly adhere to the following constraints: Multi-server usage is permitted, model weight quantization is allowed but cannot go below 4-bit precision, and model pruning is strictly prohibited, and reusing KV cache from previous inference runs to reduce computation is not allowed.

3. **DeepSeek R1-distill 32B download:** https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-32B
4. **Reference:** Participants may use vLLM on CPU (https://vllm.hyper.ai/docs/getting-started/installation-with-cpu/) as a reference implementation. Alternatively, you are free to utilize other frameworks or develop an inference system from scratch.

VII. Group Competition (5 points):

1. **Application will be posted within 5 days following the group formation.**
2. **Platform requirement:** The Runtime power consumption must not exceed 4 KW. Non-compliance will lead to immediate task disqualification
3. **Goal:** During the ASC25 finals, the competition committee will unveil multiple benchmark workloads, challenging participants to minimize execution time while ensuring that all submitted results pass correctness verification. Key restriction: Code modifications related to method parameters are strictly prohibited.

Evaluation Methods:

Applications		Points	Evaluation method
Group competition		**5**	$\forall S_j \in \{S_1, \dots, S_N\}$, where N is the number of workloads, S_j is the full score of the j^{th} workload, the score P_{se} of each group will be given as: $$P_{se} = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{T_{j\ min}}{T_{ij}} * S_j\right)$$ Where M is the number of teams in that group, T_{ij} is the runtime of the j^{th} workload achieved by the i^{th} team within that group, and $T_{j\ min}$ is the minimum among all the participating teams. A zero point will give for the workload which doesn't pass the correctness checking.
Performance Optimization (90 points)	**HPL**	**6.5**	Let S_i be the actual performance of each team in which S_{max} is the maximum of all teams, the score P_1 will be given as: $$P_1 = \left({S_i}/{S_{max}}\right) * P_C + P_C$$ Where $P_C = 3.25$ if the team gets correct result, or $P_C = 0$ if the team gets no results or invalid result.

	HPCG	**6.5**	P_2 is calculated in the same way as P_1 in HPL.
	AlphaFold3 Inference	**18**	$\forall S_i \in \{S_1, \dots, S_N\}$, where N is the number of workloads, S_i is the full score of the i^{th} workload, the score P_3 will be given as: $$P_3 = \sum_{i=1}^{N} (\frac{T_{i\ min}}{T_i} * P_s + P_s)$$ Where T_i is the runtime of the i^{th} workload, and $T_{i\ min}$ is the minimum among all the participating teams. Where $P_S = \frac{S_i}{2}$ if the team gets correct result, or $P_S = 0$ if the team gets no results or invalid result.
	RNA m^5C	**18**	P_4 is calculated in the same way as P_3
	Geant4	**18**	P_5 is calculated in the same way as P_3
	DeepSeek Inference	**18**	P_6 is calculated in the same way as P_3
	Total Points		$$P = P_{se} + \sum_{i=1}^{6} P_i$$

Team Presentation Requirements & Evaluation (10 points)

1. Presentation Format

- Each team must present their results using PowerPoint (PPT) slides.
- The presentation order will be determined by a random draw.
- Both the slides and oral delivery must be conducted in English, with up to two student representatives presenting per team.

2. Presentation Time & Jury Questions

- Presentations are limited to 7 minutes. Exceeding this duration will result in point deductions.
- Following each presentation, judges will conduct a 3-minute Q&A session to assess the team's technical understanding and problem-solving skills.

3. Presentations Evaluation & Scoring Criteria

Presentations will be evaluated and scored by the ASC25 Evaluation Committee using a 10-point scale based on:

- Clarity and effectiveness of communication
- Technical accuracy and depth
- Adherence to competition guidelines
- Innovation in optimizing workloads

4. Advisor Participation

Team advisors may observe their team's presentation session. However, advisors are strictly prohibited from interacting with judges, participants, or audiovisual equipment during the event.