

Comparing Multi-Agent LLM Negotiation Protocols for Aerodynamic-Acoustic Trade-Off Design

2026-07-14

ChinaRxiv

View the original and related papers at
<https://chinaxiv.org/items/chinaxiv-202607.00263>

Source: ChinaXiv. Machine translation. Verify with the original.

Abstract

Large language model (LLM) multi-agent systems can now autonomously operate engineering physics solvers, but existing studies treat inter-agent interaction as an implementation detail—adopting a single fixed coordination mechanism without examining it as an independent variable. This paper investigates whether the same team of LLM agents, equipped with the same physics tools, produces systematically different design outcomes under different negotiation protocols. We construct three role-specific agents—an aerodynamic engineer, an acoustic engineer, and a chief engineer—sharing a CST→Xfoil→BPM physics pipeline to perform aerodynamic-acoustic airfoil trade-off design. Each agent adopts a modular “base prompt + protocol-specific addendum” architecture, ensuring that agent capabilities remain identical across protocols and that only the interaction rules differ. We implement three structurally distinct protocols: sequential deliberation (turn-taking debate with chief-engineer adjudication), constraint-driven coordination (round-by-round dynamic tightening of the OASPL constraint), and parallel comparison (two-branch independent exploration, intra-branch quality control, and best-selection at the merging stage). Controlled experiments using DeepSeek V4 Pro as the common LLM backend and three independent runs for each protocol show that all three protocols robustly improve the lift coefficient while reducing noise relative to the NACA~0012 baseline (average $CL \sim +40\text{--}46\%$, $OASPL_{\text{\$}0.3\text{--}0.7\text{dB}}$). The result patterns exhibit systematic divergence: constraint-driven coordination achieves the most consistent noise reduction with the fewest solver calls, sequential deliberation attains the highest average CL gain per thousand Xfoil calls, and parallel comparison produces the most stable outcomes with the lowest between-run variance in OASPL. These results establish interaction architecture—how agents organize negotiation—as a first-order design dimension in multi-agent engineering systems.

Full Text

Comparing Multi-Agent LLM Negotiation Protocols for Aerodynamic-Acoustic Trade-Off Design

Zhonglu Lin*, et al.

*Shanghai Aviation Industry (Group) Co., Ltd., COMAC, Shanghai 201210, China

Abstract—Multi-agent large language model (LLM) systems can operate engineering physics solvers autonomously, but prior work treats agent interaction as an implementation detail—a single fixed coordination mechanism. This paper asks whether different negotiation protocols lead the same team of LLM agents to systematically different design outcomes. We construct three role-specialized agents—

AeroAgent, AcousticAgent, and ChiefAgent—sharing a common CST-to-Xfoil-to-BPM physics pipeline for aerodynamic-acoustic wing design. Each agent uses a modular base-plus-addon prompt architecture, ensuring identical agent capability across protocols while varying only interaction rules. We implement three structurally distinct protocols: Sequential Deliberation (turn-based debate with chief ruling), Constraint-Driven (iterative OASPL constraint tightening round by round), and Parallel Comparison (dual-branch exploration with branch-level quality gate and merge-phase selection). Controlled experiments with a single LLM backend (DeepSeek V4 Pro), using three independent runs per protocol, demonstrate that all three protocols robustly improve both lift coefficient and noise over the NACA 0012 baseline (mean CL +40-46%, OASPL -0.3 - 0.7 dB). Outcome patterns diverge systematically: Constraint-Driven yields the most consistent noise reduction with the fewest solver calls, Sequential Deliberation achieves the highest mean CL gain per Xfoil call, and Parallel Comparison produces the lowest OASPL variance across runs. These results establish interaction architecture—how agents are organized to negotiate—as a first-class design dimension in multi-agent engineering systems.

Index Terms—multi-agent systems, large language models, negotiation protocols, aerodynamic design, aero-acoustic trade-off, computational design, agent-based engineering

I. Introduction

Civil aircraft wing design confronts a structural trade-off: increasing camber raises lift but thickens the trailing-edge boundary layer, intensifying trailing-edge noise. Conventional resolution strategies fall into two categories. Multi-Disciplinary Design Optimization (MDO) reduces competing objectives to a single scalar through weighted-sum or epsilon-constraint methods—mathematically tractable but blind to the structured negotiation that real engineering teams practice. In industry, cross-disciplinary coordination relies on Integrated Product Teams and formal design reviews, where aerodynamicists and acoustics specialists negotiate trade-offs through structured interaction protocols [1]. Boeing’s 777 development program explicitly credited its “Working Together” IPT philosophy—in which the chief engineer “created constructive tension” across disciplines—for surfacing design conflicts that single-objective optimization would have suppressed [1].

Recent work has demonstrated that large language model (LLM) agents can autonomously operate engineering physics solvers. Kumar and Karniadakis [2] showed LLM agents using AeroSandbox and Xfoil for single-airfoil optimization through generator-critic loops. Xu et al. [3] built Engineering.ai, a multi-agent platform with BPM noise prediction and Xfoil aerodynamics, in which specialized agents coordinate through a fixed weighted-sum scalarization mechanism—

not through negotiation. Subsequent work extended this paradigm to MDAO frameworks [4], risk-aware set-based design [5], and turbomachinery [6]. Across this growing body of work, agent interaction is a fixed implementation choice—one coordinator, one debate structure, one turn order. No prior study treats interaction structure itself as an independent experimental variable.

In parallel, research on multi-agent debate has established that protocol design meaningfully affects interaction quality. AutoGen [7] provides the GroupChat and RoundRobin infrastructure for flexible multi-agent conversation. Marandi [8] conducted the first controlled protocol comparison, evaluating three debate structures in a macroeconomic case study with five random seeds, demonstrating that protocol design affects convergence speed, peer-referencing rate, and argument diversity. Mol-Debate [9] applies multi-agent debate to molecular design, showing that iterative generate-debate-refine loops outperform single-pass generation. These works establish that protocol design matters—but exclusively in pure text-reasoning domains. No prior work studies protocol comparison for agents operating physics-based engineering tools under aerodynamic and acoustic constraints.

We combine Marandi’s protocol-comparison methodology [8] with the engineering-agent paradigm of Kumar and Xu, asking: when the same team of LLM agents, equipped with identical physics tools, negotiates through different interaction protocols, do they converge to systematically different designs? We construct a three-agent tool-calling framework integrating CST parameterization, Xfoil aerodynamics, and BPM trailing-edge noise prediction. We implement three negotiation protocols—Sequential Deliberation, Constraint-Driven, and Parallel Comparison—each mapping to a recognizable real-world engineering decision mode. Controlled experiments with DeepSeek V4 Pro as the common LLM backend, using three independent runs per protocol, demonstrate that protocol structure produces systematic, quantifiable differences in design outcomes and resource utilization. The central finding is that interaction architecture—how agents are organized to work together—shapes design outcomes more fundamentally than individual agent capability, establishing a new axis for engineering AI research beyond agent intelligence and tool

Figure 1 diagram labels.

Physics pipeline

CST geometry → Xfoil (aerodynamics) → BPM (TE noise)

3 flight conditions: takeoff / climb / cruise (weighted)

Agent layer

call solvers

Shared base prompt + protocol-specific add-ons

AeroAgent
CST δ , Xfoil
design-space tools

AcousticAgent
BPM, compliance
OASPL limits

ChiefAgent
arbitrate, select
audit (read-only)

Negotiation protocols

Sequential
Deliberation

Constraint-
Driven

Parallel
Comparison

schedule
turn order

Fig. 1. Unified framework: physics pipeline (top), agent layer with shared base prompt and protocol-specific addons (middle), and three negotiation protocols (bottom).

quality.

II. Methodology

A. Physics Pipeline

The computational pipeline follows the standard CST-to-Xfoil-to-BPM workflow. CST/Kulfan parameterization [10] via AeroSandbox [11] represents the NACA 0012 baseline as an 8-dimensional design space (four upper-surface and four lower-surface Bernstein weights, $n_{\text{weights_per_side}} = 4$), with leading-edge weight and trailing-edge thickness held fixed at baseline-fitted values.

Xfoil [12] evaluates aerodynamics with forced transition at 5% chord on both surfaces ($x_{tr} = 0.05$) and boundary-layer data output enabled. Designs are assessed across three flight conditions: takeoff ($\alpha = 8^\circ$, $\text{Ma} = 0.18$, $\text{Re} = 4.5 \times 10^6$), climb ($\alpha = 5^\circ$, $\text{Ma} = 0.20$, $\text{Re} = 5 \times 10^6$), and cruise ($\alpha = 2^\circ$, $\text{Ma} = 0.25$, $\text{Re} = 6 \times 10^6$), with weighted scoring (0.4/0.4/0.2).

BPM trailing-edge noise prediction [13] uses the TBL-TE mechanism only (Brooks 1989 Eqs. 25-49). The Python implementation was validated against the OpenFAST Fortran TBLTE subroutine [14], achieving code-to-code agreement of $\pm 5 \times 10^{-5}$ dB per one-third-octave frequency bin. OASPL

is computed from the combined SPL spectrum without A-weighting. The reference design-space map is generated via Latin Hypercube Sampling (LHS, $n = 600$) [15], from which the Pareto front is extracted as the non-dominated subset of all evaluated candidates.

Fig. 1 illustrates the unified framework: the data flow from CST geometry through Xfoil to BPM, agent-tool assignments, and protocol control flow.

TABLE I
Agent Role Matrix.

Agent	Responsibility & Design Philosophy
AeroAgent	Proposes geometric modifications; evaluates CL, CD, δ^* . Exploration-driven; operates solvers directly.
AcousticAgent	Predicts OASPL via BPM; assesses compliance. Constraint-guarding; sets OASPL limits when authorized.
ChiefAgent	Arbitrates conflicts; selects best designs; audits results. Read-only solver access; rules by priority chain: feasibility $\succ$ compliance $\succ$ performance.

B. Agent Design

Three role-specialized agents share a common coordination mechanism: a DesignStateManager that holds all computed candidates, solver counts, and compliance records. Each agent is an independent LLM call (DeepSeek V4 Pro).

The AeroAgent proposes geometric modifications through CST delta vectors and evaluates aerodynamic performance. It operates primarily through design-space exploration tools: evaluating individual designs, batch-evaluating design sets, and conducting LHS global sampling. The AcousticAgent predicts noise via BPM and assesses compliance against simulated certification thresholds. In protocols where it takes an active constraint-setting role, it writes OASPL limits into shared state. The ChiefAgent arbitrates cross-disciplinary conflicts by reviewing structured data from both specialists. It selects best candidates, audits suspicious results through independent solver re-runs, and issues rulings. The ChiefAgent never directly operates solvers—it reads results through query tools.

Agent architecture follows a base-plus-addon pattern: a shared baseline system prompt defines each agent’s identity, engineering principles, and common tools. Protocol-specific behavior—interaction order, constraint-setting authority, and termination rules—is injected through a compact addon prompt fragment appended at agent creation. This ensures that agent capability is held constant

across protocols while only interaction rules vary. Table I summarizes each agent's role.

C. Three Negotiation Protocols

All three protocols share a common warmup phase: AeroAgent performs global design-space exploration through LHS sampling, AcousticAgent evaluates all candidates acoustically, and ChiefAgent issues initial directional guidance. The loop phase then diverges, as illustrated in Fig. 2.

Sequential Deliberation mirrors engineering design review meetings. In each round, AeroAgent proposes and evaluates modifications, AcousticAgent critiques noise implications, and ChiefAgent reviews both reports and rules. Each agent sees the full conversation history. The ChiefAgent's ruling follows an explicit priority chain: physical feasibility (Xfoil convergence, geometric validity) precedes constraint compliance, which

Figure 2 diagram text

- Shared warmup: *LHS global exploration + initial Chief guidance*
- **Sequential Deliberation:** Aero $\rightarrow$ Acoustic $\rightarrow$ Chief arbitrate $\rightarrow$ accept: select best. Feedback loop: Chief arbitrate $\rightarrow$ Aero (*iterate*).
- **Constraint-Driven:** Acoustic: set/tighten OASPL limit $\rightarrow$ Aero: optimize $\rightarrow$ Acoustic: verify $\rightarrow$ Chief arbitrate $\rightarrow$ accept: select best. Feedback loops: Acoustic verify $\rightarrow$ Aero optimize (*constraint*); Chief arbitrate $\rightarrow$ Acoustic set/tighten OASPL limit (*iterate*).
- **Parallel Comparison:** Chief: 2 directives $\rightarrow$ Branch A Ae $\cdot$ Ac $\cdot$ MinorChief and Branch B Ae $\cdot$ Ac $\cdot$ MinorChief (*isolated threads*) $\rightarrow$ ChiefMerge Pareto + compliance $\rightarrow$ final selection.

Fig. 2. Protocol control flow for Sequential Deliberation (left), Constraint-Driven (center), and Parallel Comparison (right).

precedes performance metrics. Termination occurs when the ChiefAgent explicitly signals acceptance.

Constraint-Driven implements an iterative tightening mechanism rather than a single compliance gate. Each round begins with AcousticAgent setting a progressively stricter OASPL target (round 1: baseline OASPL, round 2: baseline−1 dB, round 3+: baseline−2 dB). AeroAgent optimizes aerodynamic performance within the current constraint. AcousticAgent then verifies compliance. If AeroAgent fails to meet the constraint after consecutive attempts, the constraint relaxes by 1 dB to prevent deadlock. ChiefAgent reviews the process and decides whether to accept or initiate another tightening round. This protocol structures the design trajectory as a progressive ratchet toward lower-noise solutions.

Parallel Comparison mirrors engineering trade studies. ChiefAgent first provides two contrasting design directives (e.g., high-lift priority versus low-noise

priority). Two independent teams, each an [AeroAgent, AcousticAgent, MinorChief] RoundRobin sequence, execute in parallel with fully isolated communication threads. Each branch's MinorChief selects the best design within its branch and signals completion. ChiefMerge then compares both branches' winners through Pareto-frontier analysis and compliance data, making a final selection. This protocol prioritizes divergent exploration and structured comparison over iterative refinement.

TABLE II
Protocol comparison—chief-selected designs (mean $\pm$ SD, $n = 3$).

Protocol	CL	CD	CL/CD	OASPL	Δ CL	Δ OASPL	msgs	Xfoil	BPM
NACA 0012	0.651	0.00935	69.6	82.67	—	—	—	—	—
Seq.	0.950 $\pm$ 0.108	0.00909	105.5	82.36	+41	12	2775		
Delib.	$br > \pm 0.80$ +								
	46.0 $\pm$ 0.063 0.00955 95.1 81.94 <								
	$br > \pm 0.31$ +								
	39.5 $\pm$ 0.041 0.00992 94.8 81.98 <								
	$br > \pm \$0.11$								

III. Results and Analysis

A. Experimental Design

Three protocols share a single LLM backend (DeepSeek V4 Pro) and three flight conditions (takeoff, climb, cruise). The NACA 0012 symmetric airfoil serves as the baseline (weighted $CL = 0.651$, $CD = 0.00935$, $OASPL = 82.67$ dB). Simulated compliance thresholds follow Part 36 Stage 3 approximations: takeoff ≤ 87 , climb ≤ 86 , cruise ≤ 86 dB. Each protocol was executed in three independent runs; results are reported as mean $\pm$ standard deviation.

B. Design Outcomes and Resource Utilization

Table II reports the Chief-selected design from each protocol against the NACA 0012 baseline. All three protocols consistently produce designs that simultaneously improve both lift coefficient and OASPL over the baseline across independent runs.

Sequential Deliberation achieves the highest mean lift improvement (+46.0%) but exhibits the largest run-to-run variance in both CL (± 0.108) and OASPL (± 0.80 dB). Its iterative debate strategy favors CL designs ($CL = 1.05$), while others favor conservative low- CD designs ($CL = 0.84$, $CD = 0.00919$).

Constraint-Driven yields the most consistent OASPL reduction (-0.73 dB, SD = $\pm \$0.31$ dB) with the fewest solver calls (mean 2793 Xfoil, 1937 BPM). The constraint ratchet systematically narrows the feasible design space. Its mean CL

(+39.5%) is slightly lower than the other protocols, consistent with the ratchet mechanism prioritizing constraint satisfaction over aggressive lift maximization.

Parallel Comparison produces the lowest OASPL variance across runs ($\pm \$0.11$ dB), reflecting the structured merge-phase selection that normalizes branch-level differences. It achieves competitive CL improvement (+44.4%) but at the highest solver cost (mean 4112 Xfoil calls), reflecting the inherent duplication of dual-branch LHS exploration.

Fig. 3 visualizes representative results: the pooled LHS-explored design cloud (gray), the extracted Pareto front (red), and per-protocol chief-selected endpoints (colored markers with error bars). Constraint-Driven endpoints cluster tightly in the low-OASPL region. Sequential Deliberation endpoints span a wider Pareto segment. Parallel Comparison endpoints reflect both branches' exploration before merge-phase selection.

Fig. 3 visual text.

Vertical axis: Weighted C_L (higher is better).

Horizontal axis: Weighted OASPL [dB] ($\leftarrow$ lower is quieter).

Legend: LHS-explored designs; Pareto front; NACA 0012 baseline; Sequential Deliberation; Constraint-Driven; Parallel Comparison.

Fig. 3. Design space: pooled LHS-explored designs (gray) across the three runs per protocol, with the extracted Pareto front (red curve) and per-protocol chief-selected endpoints (mean $\pm$ SD, $n = 3$).

TABLE III
NORMALIZED RESOURCE EFFICIENCY (MEAN OF $n = 3$ RUNS).

Protocol	C_L gain / 10^3 Xfoil	Δ OASPL / 10^3 BPM
Seq. Delib.	0.091	0.14 dB
Constr.-Driven	0.092	0.38 dB
Parallel Comp.	0.070	0.25 dB

To compare protocol efficiency while accounting for differing resource footprints, we compute normalized efficiency ratios (Table III).

Fig. 4 contrasts the symmetric NACA 0012 baseline with each protocol's Pareto-best section. All three protocols converge on cambered profiles—the classic mechanism for raising lift—while preserving thickness, lifting C_L by 48–50% and L/D from 70 to 96–100. The accompanying noise reductions keep every design within Part 36 Stage 3 limits, illustrating that the negotiated aerodynamic gains do not come at the cost of acoustic certification margin.

Constraint-Driven achieves the highest OASPL reduction per BPM call (0.38 dB / 10^3 BPM), more than double Sequential Deliberation's acoustic efficiency. Its constraint ratchet directs BPM evaluations toward designs likely to satisfy

tightened limits. Sequential Deliberation achieves high C_L gain per Xfoil call (0.091) but the lowest acoustic efficiency—its debate structure invests solver effort broadly rather than targeting specific constraint boundaries. Parallel Comparison's C_L efficiency is lowest (0.070), directly reflecting the overhead of dual-branch LHS exploration. Message counts converge to similar means across protocols (133–135), confirming that protocol structure, not communication volume, drives outcome

Fig. 4 visual text.

Airfoil-section plot horizontal axis: x/c .

Legend: NACA 0012; Seq. Delib.; Constr.; Parallel.

Airfoil	C_L	L/D	OASPL	ΔC_L	ΔdB
Baseline NACA 0012	0.651	69.6	82.67	—	—
Sequential Delib.	0.966	99.7	82.47	+49%	-0.20
Constraint-Driven	0.975	97.5	82.30	+50%	-0.37
Parallel Comp.	0.978	96.2	82.10	+50%	-0.57

All designs within Part 36 Stage 3 limits (takeoff ≤ 87 , climb/cruise ≤ 86 dB).

Fig. 4. Baseline vs. per-protocol Pareto-best airfoil sections (real CST geometry) with three-condition weighted performance. Optimized sections add camber while preserving thickness; ΔC_L and ΔdB are relative to the NACA 0012 baseline.

differences.

C. Qualitative Observations

Agent communication density adapts to protocol context. Sequential Deliberation elicits richer argumentation—agents reference prior-round data and build on previous proposals. Constraint-Driven produces concise, compliance-focused exchanges centered on constraint values and verification results. Parallel Comparison surfaces branch-level competition dynamics, with MinorChief agents independently selecting branch winners before the merge-phase comparison.

The constraint ratchet in Constraint-Driven is empirically well-calibrated: the baseline OASPL of approximately 82 dB is tightened in roughly 1 dB increments. The 1 dB relaxation mechanism activates when AeroAgent fails consecutive compliance attempts, preventing deadlock while maintaining overall downward pressure on OASPL. AeroAgent occasionally required multiple attempts to find feasible CST modifications under tight OASPL constraints in Constraint-Driven rounds 2–3, where the feasible design sub-space narrows considerably.

Across all runs, no fabricated numerical values were detected. The ChiefAgent correctly rejects infeasible designs—for instance, when AeroAgent proposes

CST modifications with wrong-sign Bernstein weights causing geometric self-intersection and Xfoil non-convergence—based on the convergence status field in the structured data layer.

IV. Discussion

The experiments demonstrate that how agents are organized to negotiate shapes design outcomes in ways not reducible to individual agent capability. With identical agents, identical tools, and an identical physics pipeline, three structurally distinct protocols produce systematically different design locations on the Pareto front, with different resource utilization patterns. This echoes a principle observed in engineering

practice: the Boeing 777's IPT structure and formal review format shaped which designs emerged [1], not merely how efficiently they were produced.

Sequential Deliberation emphasizes iterative refinement at moderate resource cost, yielding the highest mean CL but with the greatest run-to-run variability. Constraint-Driven enforces progressive discipline through its ratchet mechanism, achieving the most consistent OASPL reduction at the lowest solver cost. Parallel Comparison maximizes exploration diversity at the cost of computational duplication, producing the most stable OASPL outcomes across runs through its structured merge-phase selection. No protocol is universally optimal—the results suggest a protocol-to-task fit.

This work does not compete with MDO. The Pareto front describes what is physically achievable; the protocols show how different organizational structures select which subset of achievable designs an agent team converges toward. The study is scoped to a single LLM backend and 2D airfoil geometry; cross-model and cross-domain generalization remain for future investigation.

V. Conclusion

We constructed a three-agent tool-calling framework for aerodynamic-acoustic wing design and implemented three negotiation protocols—Sequential Deliberation, Constraint-Driven, and Parallel Comparison—each formalizing a distinct real-world engineering decision mode. Controlled experiments with a single LLM backend, using three independent runs per protocol, demonstrate that protocol structure systematically shapes design outcomes and resource utilization. Constraint-Driven achieves the most consistent noise reduction with the fewest solver calls through its iterative tightening mechanism. Sequential Deliberation yields the highest mean lift improvement through structured debate, with greater outcome variability. Parallel Comparison produces the most stable OASPL outcomes through dual-branch exploration with quality-gated selection. All three protocols robustly improve both lift and noise over the NACA 0012 baseline across independent runs. These results establish interaction architecture—how agents are organized to negotiate—as a first-class design dimension in

multi-agent engineering systems, complementary to agent capability and tool quality.

References

- [1] ICAS, “Innovative aspects of the Boeing 777 development program,” in *Proc. 20th Congress of the International Council of the Aeronautical Sciences (ICAS)*, Sorrento, Italy, 1996, paper ICAS-96-0.4.
- [2] V. Kumar and G. E. Karniadakis, “Toward autonomous engineering design: A knowledge-guided multi-agent framework,” 2025. [Online]. Available: <https://arxiv.org/abs/2511.03179>
- [3] R. Xu, Y. Qi, J. Feng, and X. Chu, “Engineering.ai: A platform for teams of AI engineers in computational design,” 2025. [Online]. Available: <https://arxiv.org/abs/2511.00122>
- [4] C. H. Lee, G. Cinar, and J. R. R. A. Martins, “Aerodynamic design and optimization via a specialized agentic generative AI framework,” 2025. [Online]. Available: <https://www.gokcincinar.com/publication/pp-2025-agenticframework/>
- [5] V. Kumar and G. E. Karniadakis, “Agentic risk-aware set-based engineering design,” 2026. [Online]. Available: <https://arxiv.org/abs/2604.16687>
- [6] P. Du, Y. Wu *et al.*, “TurboAgent: An LLM-driven autonomous multi-agent framework for turbomachinery aerodynamic design,” 2026. [Online]. Available: <https://arxiv.org/abs/2604.06747>
- [7] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang, “AutoGen: Enabling next-gen LLM applications via multi-agent conversation,” 2023. [Online]. Available: <https://arxiv.org/abs/2308.08155>
- [8] R. Z. Marandi, “The impact of multi-agent debate protocols on debate quality: A controlled case study,” 2026. [Online]. Available: <https://arxiv.org/abs/2603.28813>
- [9] W. Zhang, X.-Y. Wei, and Q. Li, “Mol-Debate: Multi-agent debate improves structural reasoning in molecular design,” 2026. [Online]. Available: <https://arxiv.org/abs/2604.20254>
- [10] B. M. Kulfan, “Universal parametric geometry representation method,” *Journal of Aircraft*, vol. 45, no. 1, pp. 142–158, 2008.
- [11] P. D. Sharpe, “AeroSandbox: A differentiable framework for aircraft design optimization,” M.Eng. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2021. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/140023>
- [12] M. Drela, “XFOIL: An analysis and design system for low Reynolds number airfoils,” in *Low Reynolds Number Aerodynamics*, T. J. Mueller, Ed. Berlin,

Germany: Springer, 1989, pp. 1-12.

[13] T. F. Brooks, D. S. Pope, and M. A. Marcolini, “Airfoil self-noise and prediction,” NASA, Washington, DC, USA, Tech. Rep. NASA RP-1218, 1989.

[14] P. Moriarty, G. Guidati, and P. Migliore, “Recent improvement of a semi-empirical aeroacoustic prediction code for wind turbines,” National Renewable Energy Laboratory, Golden, CO, USA, Tech. Rep. NREL/TP-500-34478, 2003.

[15] M. D. McKay, R. J. Beckman, and W. J. Conover, “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,” *Technometrics*, vol. 21, no. 2, pp. 239-245, 1979.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv. Machine translation. Verify with the original.