

ORIGINAL RESEARCH

Open access

On the Conditions under Which Sequential Learning Fails: Catastrophic Forgetting in Compositionally Complex Alloys

Lucas Meyer^{1*}, Anna Schmid², Stefan Braun¹

Abstract

Sequential learning is increasingly invoked in compositionally complex alloy modeling because the combinatorial scale of alloy design space, coupled with the cost of density-functional theory calculations and the staged nature of data acquisition, makes exhaustive one-shot training impractical. Yet under these conditions, neural networks are prone to catastrophic forgetting, whereby adaptation to newly introduced compositions, phases, or thermodynamic regimes degrades performance on previously learned ones. This article examines the conditions under which such failure emerges in compositionally complex alloys and shows that forgetting is not incidental but structurally induced by the interaction of shared representations, non-linear property landscapes, high compositional dimensionality, and sparse task-specific data. The analysis identifies four recurrent failure modes—elemental forgetting, concentration range collapse, phase space amnesia, and temperature-induced overwrite—and links them to underlying mechanisms of parameter interference, representation drift, output distribution shift, and gradient conflict. It further delineates the conditions that intensify these dynamics, including high compositional similarity, abrupt task transitions, rehearsal-free updates, and unstable optimization settings. By integrating detection criteria with mitigation strategies such as elastic weight consolidation, memory replay, modular architectures, and curriculum sequencing, the study provides a practical and conceptual framework for continual learning in alloy discovery. The central argument is that robust sequential learning in compositionally complex alloys requires forgetting-aware design at the levels of evaluation, architecture, and training protocol if materials machine learning is to support reliable lifelong modeling across expanding compositional spaces.

Keywords Continual learning, High-entropy alloys, Catastrophic forgetting, Sequential learning, Compositionally complex alloys, Stability-plasticity dilemma

*Correspondence:

Lucas Meyer

lucas.meyer@gmail.com

¹ Department of Intelligent Materials Engineering, Faculty of Engineering, ETH Zurich, Zurich, Switzerland

² Department of Computational Materials Analytics, Faculty of Science, University of Bern, Bern, Switzerland

Introduction

Compositionally complex alloys (CCAs), including high-entropy alloys (HEAs), occupy an exceptionally large compositional hyperspace in which the number of conceivable alloy systems increases rapidly with the inclusion of additional principal elements. This vast search space has motivated machine-learning-enabled alloy discovery and property-targeted design strategies [1, 2]. One-shot learning, in which a single model is trained simultaneously on all available compositions, phases, and conditions, is often impractical. The underlying reasons are threefold: first, sufficiently dense DFT or experimental datasets across the full compositional space are expensive to generate; second,

new elements, concentrations, or metastable phases are commonly characterized incrementally as synthesis and simulation campaigns progress; and third, useful alloy models must often incorporate emerging data without being retrained from scratch each time. Sequential learning therefore offers a pragmatic route for materials discovery workflows in which data arrive in stages rather than all at once [3].

Yet this incremental accumulation of knowledge collides with a well-documented limitation of neural networks. McCloskey and Cohen identified catastrophic interference as a fundamental problem in connectionist networks trained sequentially [4], and Kirkpatrick *et al.* later formalized one influential solution, elastic weight consolidation, for deep neural networks [5]. More broadly, continual learning research has shown that neural models must continually negotiate the stability-plasticity dilemma: the need to acquire new knowledge while preserving previously learned knowledge [6, 7]. In the materials domain, this problem becomes especially important because machine-learning interatomic potentials and graph-based neural models often reuse shared representations across many related chemical environments [8–10]. Lifelong machine-learning-potential frameworks have explicitly recognized this issue by treating continued model adaptation as a central design problem rather than a peripheral training detail [11].

The challenge is particularly acute in high-entropy and compositionally complex alloys. Machine learning has been applied to HEA property prediction, catalyst design, microstructure simulation, and broader alloy-design tasks [12–15]. However, the same features that make CCAs scientifically attractive also make them difficult for sequential neural learning: overlapping local atomic motifs, sparse data, non-linear phase stability, and high-dimensional chemical interactions. The central problem addressed in this article is therefore clear: sequential learning in CCAs, where new elements, compositions, phases, or temperatures are added incrementally, can induce catastrophic forgetting. As a result, previously learned knowledge of formation energies, elastic properties, phase stability, or local atomic interactions may be rapidly overwritten. Because such failures can compromise downstream deployment, forgetting-aware evaluation should be treated as part of a broader verification-aware approach to neural-network use in scientific modeling [16].

Figure 1 presents a hierarchical failure architecture showing how the structural properties of compositionally complex alloys propagate through forgetting mechanisms, trigger specific failure modes, and necessitate detection and mitigation in sequential learning.

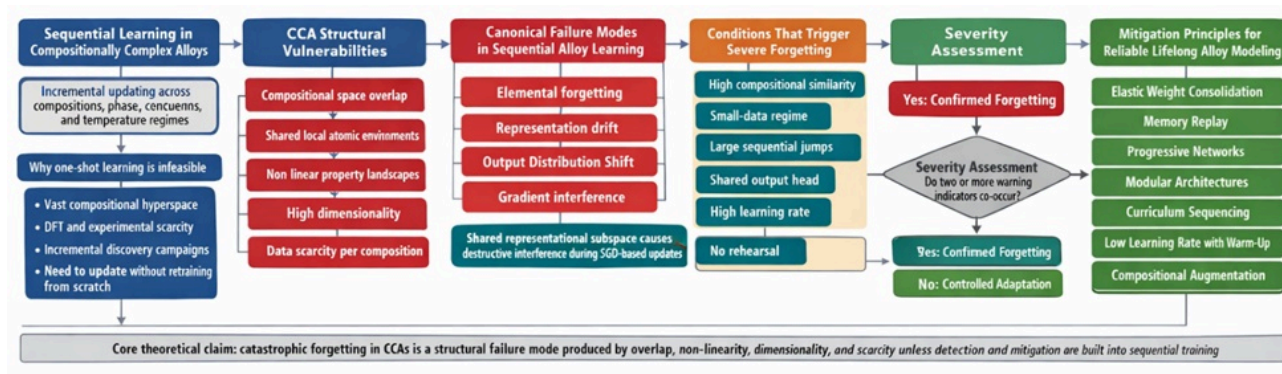


Figure 1. Hierarchical Failure Architecture of Sequential Learning Breakdown in Compositionally Complex Alloys

The problem can be summarized as follows. CCAs inhabit an enormous compositional space. One-shot learning is frequently infeasible because of computational cost, data availability, and experimental constraints. Sequential learning is therefore a natural alternative: models are first trained on simpler systems, such as binary or ternary alloys, and then updated with data from higher-order compositions, new phases, or new temperature regimes. However, neural networks can suffer from catastrophic forgetting, meaning that learning new information degrades performance on previously learned distributions [4, 5]. In CCAs, this failure is amplified by the physical and statistical structure of the alloy design space.

Catastrophic Forgetting in Sequential Learning

A precise conceptualization of catastrophic forgetting is essential when situating the phenomenon within materials machine learning. In this context, catastrophic forgetting refers to the pronounced degradation in predictive performance on previously encountered data distributions when a neural model is updated sequentially on new tasks without sufficient protection of earlier knowledge. In alloy modeling, a “task” may correspond to a compositional family, a concentration interval, a phase class, or a thermodynamic condition. For example, a model may first learn binary Co–Cr structures, then ternary Co–Cr–Ni structures, and later quaternary or quinary CCAs. The model is expected to accumulate knowledge, but each new training stage modifies the same parameter set that previously encoded earlier alloy behavior.

The notion of a forgetting rate provides a useful analytical lens. A forgetting rate can be defined as the increase in prediction error on an earlier validation set after training on a later task. If a model achieves low error on a binary alloy dataset but that error rises sharply after training on a quinary dataset, the model has not merely failed to generalize; it has lost prior competence. This distinction matters because final-task accuracy alone can conceal severe degradation on earlier tasks.

Standard stochastic gradient descent is a central contributor to this problem. During sequential training, mini-batches are drawn primarily or exclusively from the current distribution. The optimizer therefore updates parameters in directions that reduce current-task loss, even when those directions increase the loss on previous tasks. Some architectural properties, including network width and capacity, can influence the severity of forgetting [17], but capacity alone does not solve the problem. Unless previous data, parameter-importance constraints, or task-specific structures are included in the training process, the model has no direct incentive to preserve earlier representations.

The consequences are especially severe in alloy modeling because local atomic environments recur across compositions. A neural network trained first on CoCr structures may learn embeddings for FCC-like local environments. Subsequent training on CoCrFeNi or CoCrFeMnNi data can perturb those embeddings, causing earlier binary or quaternary predictions to drift. Rohr *et al.* showed that sequential learning can accelerate materials discovery, but their benchmark also highlights the need to evaluate learning trajectories rather than only final outcomes [3]. Lifelong machine-learning-potential work similarly emphasizes that models adapted over time need explicit strategies to retain earlier competence [11].

A conceptual diagram of catastrophic forgetting can be imagined as a two-axis plot with training epochs on the x-axis and prediction error on the y-axis. One curve represents the previous task, such as binary alloy properties, and another represents the current task, such as ternary or quinary alloy properties. During training on the first task, error on that task falls. When training switches to the second task, error on the new task falls, but error on the first task rises. If the rise is abrupt and sustained, the curve displays the classic signature of catastrophic forgetting.

The alloy setting differs from many benchmark classification problems because tasks are not fully independent. In image classification, task categories may be semantically distant. In CCA modeling, however, tasks often share local environments, descriptors, elements, and structural motifs. This representational overlap means that forgetting is not simply caused by task dissimilarity. It can also arise from excessive similarity, because chemically related tasks compete for the same representational subspace. Catastrophic forgetting in sequential alloy learning is therefore a structural outcome of how gradient-based optimization interacts with entangled compositional representations.

Why Compositionally Complex Alloys Are a Worst-Case Scenario

Compositionally complex alloys expose a particularly acute manifestation of catastrophic forgetting because several sources of interference occur simultaneously. The difficulty does not arise from a single factor in isolation. Instead, it

emerges from the interaction of overlapping local atomic environments, non-linear property landscapes, high compositional dimensionality, and sparse data.

First, CCAs exhibit strong compositional-space overlap. Similar short-range motifs, including FCC, BCC, and chemically mixed coordination environments, may appear in different alloy families. A neural network potential trained on one CCA family may therefore reuse internal descriptors when learning another. This reuse is useful for transfer, but it also creates vulnerability. Updates that improve the representation of a motif in one chemical context can distort the representation of the same or similar motif in another context.

Second, CCA property landscapes are highly non-linear. Formation energies, elastic constants, phase stability, and defect energetics do not necessarily vary smoothly with composition. Small changes in elemental fractions can cause disproportionate changes in phase preference or mechanical response. Machine-learning phase-prediction studies for high-entropy alloys demonstrate that model performance is sensitive to descriptor selection, training distribution, and the structure of the composition space [18, 19]. In sequential learning, these non-linearities make it risky to assume that knowledge learned from one region of the alloy space can be smoothly adjusted to another without damaging earlier mappings.

Third, the dimensionality of CCA design space expands rapidly as additional principal elements are included. A five-element system contains many more possible local chemical neighborhoods than a binary or ternary system. Each new element introduces additional interaction terms that must be represented by the same hidden layers unless the architecture explicitly separates them. Even expressive graph neural networks and equivariant potentials remain susceptible to interference when sequential updates force a shared model to accommodate new graph topologies, element types, or coordination environments [10].

Fourth, many CCA datasets are sparse. DFT calculations are expensive, and experimental measurements are unevenly distributed across compositions and processing histories. Models trained on sparse early tasks may form brittle representations that work only in narrow regions of composition space. When new data arrive, these fragile representations can be overwritten quickly. Sparse data therefore magnify the instability produced by sequential optimization.

These four factors make CCAs more vulnerable to forgetting than simpler elemental or binary systems. In a pure element or a densely sampled binary alloy, local environments may be more distinct, property trends more regular, and training data more abundant. In a CCA, by contrast, the model is asked to maintain many partially overlapping, non-linear, and sparsely constrained mappings at once. Sequential learning under these conditions becomes a worst-case scenario for catastrophic forgetting.

Mechanisms of Forgetting in Alloy ML

Catastrophic forgetting in sequentially trained alloy models can be understood through four interacting mechanisms: parameter interference, representation drift, output-distribution shift, and gradient interference.

Parameter interference is the most immediate mechanism. Neural-network parameters are usually shared across all tasks. A single weight matrix may contribute to predictions for binary, ternary, quaternary, and quinary systems. When new-task training updates those weights, it may move them away from values that were important for earlier tasks. Elastic weight consolidation was proposed precisely to limit such destructive movement by penalizing changes to parameters that are important for previous tasks [5]. In the absence of such protection, shared parameters can be repurposed by the current task at the expense of prior tasks.

Representation drift occurs when hidden-layer embeddings change after sequential updates. In alloy models, hidden layers often encode local atomic neighborhoods. If a model has learned an embedding for a Co–Cr–Ni environment, the

introduction of Mn or Fe may cause the embedding space to be reorganized. Even if the output layer remains unchanged, predictions can degrade because the features feeding that output layer no longer have the same meaning. Representation drift is especially damaging in CCAs because related environments are close enough to interfere but different enough to require distinction.

Output-distribution shift occurs when the range or shape of target properties changes across tasks. A model trained on binary alloys may learn formation energies within a relatively narrow interval. A later quinary CCA dataset may contain a broader or differently distributed energy range. During fine-tuning, the model may recalibrate its output layer to fit the new range, distorting earlier predictions. This mechanism is especially relevant when a single shared output head is used for all compositions, phases, or temperature regimes.

Gradient interference occurs when gradients for the current task conflict with gradients that would preserve earlier performance. If the update direction that reduces new-task loss increases old-task loss, the model experiences negative backward transfer. This mechanism links the geometry of optimization directly to catastrophic forgetting. In CCAs, gradient conflict is likely because new tasks often combine familiar motifs with new elements, phases, or property ranges. The model is therefore pulled toward parameter regions that satisfy the current task but destabilize earlier ones.

These mechanisms reinforce one another. Parameter interference initiates weight displacement. Representation drift changes the internal meaning of local-environment embeddings. Output-distribution shift corrupts calibration. Gradient interference accelerates all three processes during optimization. The result is often not slow degradation but an abrupt collapse of prior-task performance once the new task begins to dominate training.

Table 1. Structural Mapping of Sequential Learning Failure in Compositionally Complex Alloys

Analytical layer	Core construct	Definition in this manuscript	CCA-specific manifestation	Primary consequence for model behavior
Problem foundation	Sequential learning	Incremental training on compositional subsets presented over time	Binary → ternary → quaternary → quinary alloy progression; phase or temperature expansion	Creates practical route for modeling vast alloy spaces without exhaustive one-shot retraining
System vulnerability	Compositional space overlap	Different alloy systems share partially similar local environments	FCC-like or chemically similar neighborhoods recur across nominally distinct CCAs	Encourages entangled internal representations
System vulnerability	Non-linear property landscape	Small composition changes induce disproportionately large property changes	Formation energy, elastic behavior, or phase stability vary non-monotonically	Limits safe parameter sharing across adjacent tasks
System vulnerability	High dimensionality	Addition of principal elements expands interaction space combinatorially	5+ element systems introduce many new local interaction terms	Increases probability of interference during updates
System vulnerability	Data scarcity per composition	Each local region of compositional space is sparsely sampled	Limited DFT structures or experimental observations for many compositions	Produces fragile, easily overwritten representations

Mechanism	Parameter interference	New-task updates move weights away from prior useful values	Shared parameters encode multiple alloy families simultaneously	Prior-task performance degrades after new optimization
Mechanism	Representation drift	Hidden-layer embeddings change after new learning	Local environment encodings shift when new elements or phases are introduced	Earlier prediction mappings become unreliable
Mechanism	Output distribution shift	Target range changes across sequential tasks	Binary and quinary systems may span different property intervals	Previously calibrated outputs become distorted
Mechanism	Gradient interference	New-task gradients increase old-task loss	Sequential alloy updates push parameters against prior optima	Negative backward transfer emerges
Failure expression	Elemental forgetting	Prior elemental knowledge collapses after element addition	CoCrFeNi accuracy deteriorates after adding Mn	Loss of performance on earlier alloy families
Failure expression	Concentration range collapse	Broad concentration behavior is lost after narrow-range fine-tuning	Model retains 20–30 at.% behavior but forgets extremes	Sharp extrapolation failure across prior range
Failure expression	Phase space amnesia	Previously learned phase relations are lost after new phase learning	Stable FCC predictions degrade after intermetallic or BCC training	Unphysical phase-ordering predictions
Failure expression	Temperature-induced overwrite	Low-temperature fidelity erodes after high-temperature updating	0 K accuracy degrades after training on finite-temperature trajectories	Multi-condition unreliability
Diagnostic implication	Forgetting as structural outcome	Forgetting arises from architecture–data interaction, not isolated bad luck	CCA properties amplify continual learning fragility	Sequential learning requires explicit forgetting-aware design

A Typology of Forgetting Failure Modes

Catastrophic forgetting in CCAs does not manifest in a single uniform way. It appears through recurring failure modes that are analytically separable but mechanistically connected.

The first mode is elemental forgetting. This occurs when the introduction of an additional element destabilizes previously learned compositional knowledge. A model trained on CoCrFeNi may perform accurately on quaternary structures, but after training on CoCrFeMnNi it may lose accuracy on the original four-element system. The new element does not simply extend the representation space; it can compress and reorganize it. Chemically similar species may become less separable in the embedding space, and previously stable predictions may drift.

The second mode is concentration range collapse. A model may initially learn property variation over a broad concentration interval, such as dilute to near-equiatomic regimes. If it is later fine-tuned only on a narrow concentration window, the model may recalibrate around that window and lose accuracy at the extremes. This failure reflects output-

distribution contraction. The most recent data slice dominates the model's response function, and earlier concentration-dependent curvature is lost.

The third mode is phase-space amnesia. A model trained on stable FCC solid solutions may later be trained on BCC structures, ordered intermetallics, or metastable phases. The gradients associated with the new structural regime can alter the learned energetic ordering of phases. As a result, structures that should remain stable may be assigned unphysical energies. This is particularly consequential in HEA systems where eutectic and multi-phase behavior can be central to alloy design [20].

The fourth mode is temperature-induced overwrite. Machine-learning interatomic potentials may be trained first on 0 K DFT-relaxed structures and later adapted to high-temperature molecular-dynamics configurations. Training on finite-temperature data can improve high-temperature performance while degrading low-temperature accuracy. In this case, the model overwrites precise low-temperature mappings with broader thermal configurations. This is not merely a data-distribution problem; it is a physical-regime problem, because the model is asked to represent different thermodynamic conditions without explicit protection of earlier knowledge.

These four failure modes give researchers a practical diagnostic vocabulary. Elemental forgetting points to loss of chemical separability. Concentration range collapse points to distributional narrowing. Phase-space amnesia points to corrupted structural ordering. Temperature-induced overwrite points to regime-specific calibration loss. All four modes arise from the same underlying mechanisms: parameter interference, representation drift, output-distribution shift, and gradient conflict.

Conditions that Trigger Catastrophic Forgetting

Catastrophic forgetting is not an unavoidable consequence of all sequential learning. Its severity depends on specific triggering conditions. In CCAs, several conditions are especially important.

High compositional similarity between successive tasks is one major trigger. If a new element resembles existing elements in atomic radius, electronegativity, valence behavior, or local coordination preference, the network may struggle to maintain distinct embeddings. For example, adding Mn to a CoCrFeNi model may introduce chemically related environments that are close to earlier ones but not identical. The feature extractor must adjust, and those adjustments can destabilize previously learned mappings.

Small-data regimes are another major trigger. When each task contains only a limited number of DFT-relaxed structures or experimental observations, the model can overfit local patterns. These overfit representations may perform well immediately after training but remain fragile. New-task updates can then overwrite them because they are not anchored by broad data support.

Large jumps in task complexity intensify forgetting. Moving directly from binary systems to quinary systems introduces a large change in both input structure and target distribution. The optimizer may respond with large parameter updates, especially if the learning rate is high. Gradual curricula are usually safer because they allow the model to build hierarchical representations before encountering the most complex tasks.

Shared output heads can also trigger forgetting. If all compositional families, phases, and temperatures are forced through a single output pathway, new-task calibration can distort old-task calibration. This is especially problematic when different tasks have different property ranges. Composition-specific, phase-specific, or temperature-specific heads can reduce this pressure.

Optimization settings matter as well. Learning rates that are suitable for one-shot training can be too aggressive for sequential training. Large step sizes accelerate parameter displacement and increase the likelihood of catastrophic

overwrite. Warm-up schedules and conservative learning rates reduce the risk by limiting abrupt movement through parameter space.

Finally, the absence of rehearsal is one of the strongest triggers. If prior data are never revisited, the model receives no corrective signal when old-task performance begins to degrade. Memory replay, synthetic replay, or compositional augmentation can provide anchors that keep earlier distributions active during new-task learning.

Detection Principles

Detection of catastrophic forgetting requires monitoring more than final-task accuracy. A sequential model can appear successful on the newest task while failing badly on earlier ones. Five detection principles are especially useful for CCA modeling.

First, performance on previous tasks should be evaluated after each new training stage. A fixed validation set from every prior compositional task should be maintained. If error on earlier tasks rises while current-task error falls, negative backward transfer has occurred. This should be measured at the level of individual compositions, not only through global averages, because average metrics can hide elemental forgetting or phase-space amnesia.

Second, hidden representations should be compared before and after new learning. Probe structures, such as fixed FCC local environments from earlier alloy systems, can be passed through the model before and after sequential updates. A sharp drop in cosine similarity between hidden activations indicates representation drift.

Third, the magnitude of parameter updates should be monitored. The normalized L2 distance between pre-update and post-update parameters can serve as an early-warning signal. Large changes in important layers suggest that the optimizer is moving the model away from earlier optima.

Fourth, a cumulative held-out validation set should span all compositions seen so far. This set acts as a global canary. If its mean error rises during training on a new task, the model is sacrificing previous knowledge for current-task performance.

Fifth, predictions should be tracked on local environments that recur across compositions. Because CCAs contain similar short-range motifs in different chemical contexts, identical or nearly identical local clusters can be used as probes. Increased prediction variance across these probes signals local-environment confusion.

When two or more detection principles flag degradation simultaneously, catastrophic forgetting should be treated as confirmed. At that point, continued training without mitigation is likely to worsen the damage.

Mitigation Principles

Effective mitigation requires adapting continual-learning methods to the structure of CCA data. The most useful strategies are regularization, rehearsal, architectural separation, curriculum design, conservative optimization, and domain-specific augmentation.

Elastic weight consolidation is a natural starting point. It penalizes changes to parameters that are estimated to be important for earlier tasks [5]. In alloy modeling, this means that parameters critical to earlier compositions or phases are stabilized during later training. EWC is especially useful when old data cannot be fully retained.

Memory replay provides a more direct anchor. A small subset of earlier structures can be stored and interleaved with new-task mini-batches. Even a modest replay buffer can remind the model of earlier distributions and reduce representation

drift. Replay is particularly effective when local environments recur across alloy families because old motifs remain visible during new training.

Architectural separation can reduce interference at the design level. Progressive networks add new model components for new tasks while freezing earlier components. Modular architectures and multi-classifier approaches provide another route: a shared backbone extracts local features, while separate heads serve different compositions, phases, or temperatures [21]. This confines some output-distribution shift to task-specific modules and protects the shared representation.

Curriculum sequencing addresses the order in which tasks are presented. Rather than jumping directly from binary to quinary systems, the model can progress from binary to ternary, quaternary, and then quinary systems. Similarly, broad concentration ranges can be learned before narrow slices, and stable phases can be learned before metastable phases. Sequential-learning benchmarks in materials discovery show why task ordering should be treated as an experimental variable rather than a convenience [3].

Conservative optimization further reduces risk. Lower learning rates, gradual warm-up, and early stopping based on cumulative validation performance can prevent abrupt parameter displacement. In sequential learning, the goal is not only to minimize current-task loss but also to preserve previous-task competence.

Domain-specific augmentation can provide a soft form of rehearsal. New-task structures can be augmented with partial substitutions or local motifs from earlier tasks, creating overlap regions that force the model to revisit prior representations. Selfless sequential learning and related approaches show how model updates can be constrained to reduce harm to earlier knowledge [22].

In practice, these methods are complementary. A robust CCA continual-learning workflow may combine EWC, replay, modular heads, curriculum ordering, and conservative optimization. The key point is that mitigation cannot be treated as optional. In high-dimensional alloy spaces, forgetting-aware design should be a default component of sequential model training.

Table 2 provides a diagnostic–mitigation matrix that specifies which detection and intervention strategies are most appropriate under different sequential-learning failure conditions in compositionally complex alloys.

Table 2. Diagnostic–Mitigation Matrix for Catastrophic Forgetting in Sequential Alloy Learning

Triggering condition	Most likely dominant mechanism(s)	Most vulnerable failure mode(s)	Recommended detection principle(s)	Priority mitigation principle(s)	Rationale for pairing
High compositional similarity between successive tasks	Representation drift; parameter interference	Elemental forgetting; phase space amnesia	Representation similarity analysis; local environment confusion	Modular architectures; EWC; compositional augmentation	Similar local environments require protection of shared representations and better compositional separation
Small-data regime	Parameter interference; over-specialization; output shift	Concentration range collapse; elemental forgetting	Cross-composition validation; parameter change magnitude	Memory replay; low learning rate with warm-up; curriculum sequencing	Sparse tasks create fragile local minima that must be anchored during updates

Large sequential jump in task complexity	Gradient interference; representation drift	Elemental forgetting; phase space amnesia; overwrite across all prior tasks	Backward transfer monitoring; parameter change magnitude	Curriculum sequencing; progressive networks; replay	Abrupt distributional expansion produces destructive gradient steps that benefit from staged adaptation
Shared output head across all tasks	Output distribution shift; parameter interference	Concentration range collapse; temperature-induced overwrite	Cross-composition validation; backward transfer monitoring	Composition-specific or phase-specific output heads; modular architectures	Shared output layers force incompatible distributions through one prediction channel
High learning rate during sequential update	Parameter interference; gradient interference	All four failure modes	Parameter change magnitude; backward transfer monitoring	Low learning rate with warm-up; EWC	Large updates rapidly dislodge prior optima
No rehearsal of previous data	Representation drift; gradient interference	Elemental forgetting; concentration range collapse; phase amnesia	Cross-composition validation; backward transfer monitoring	Memory replay; compositional augmentation	Past distributions need explicit anchors during new-task learning
New phase family introduced after stable-phase training	Gradient interference; output shift	Phase space amnesia	Cross-composition validation; local environment confusion	Replay; modular heads; progressive networks	New phase statistics can corrupt previously learned thermodynamic ordering
High-temperature data added after low-temperature training	Output distribution shift; representation drift	Temperature-induced overwrite	Backward transfer monitoring on 0 K validation; parameter change magnitude	Modular heads; replay; reduced learning rate	Sequential thermal broadening can destroy precise low-temperature calibration

Implications for Continual Learning in Materials

Recognizing catastrophic forgetting as a structural property of sequential learning in CCAs reframes it from an occasional nuisance to a central constraint on model design and evaluation. For researchers developing models, final accuracy is not enough. A model that performs well on the newest task but poorly on earlier tasks has not learned cumulatively. Reports of sequential alloy learning should therefore include backward transfer, forgetting rates, cumulative validation curves, and task-resolved error.

This reorientation also changes benchmark design. Static datasets are insufficient for evaluating lifelong alloy models. Benchmarks should encode task order, compositional similarity, data scarcity, phase transitions, and temperature shifts. A useful benchmark should make forgetting observable rather than accidentally hidden. Sequential-learning protocols should therefore report not only what data were used but also the order in which they were presented.

Software infrastructure must also adapt. Materials machine-learning frameworks often assume one-shot training. Continual alloy modeling requires native support for replay buffers, parameter-importance estimation, modular heads, progressive components, cumulative validation, and forgetting metrics. These features should not be treated as optional add-ons. They are necessary for reliable deployment in expanding alloy spaces.

The broader implication is conceptual. Materials machine learning has often operated episodically: train a model on a fixed dataset, evaluate it, and publish the final score. CCAs demand a cumulative paradigm. Models must remain reliable as the design space expands. This is especially important for high-entropy and refractory high-entropy alloy discovery, where the search space is large and only sparsely explored [23, 24]. Under a forgetting-aware paradigm, sequential learning can become a reliable pathway for scalable discovery rather than a hidden source of model fragility.

Catastrophic forgetting should be understood as a central limitation of sequential learning in compositionally complex alloys rather than a peripheral training artifact. CCA modeling combines overlapping local environments, sparse data, expanding compositional dimensionality, and strongly non-linear structure-property relationships. These conditions make sequential updates likely to destabilize previously learned knowledge unless explicit safeguards are introduced.

This article has shown that forgetting in CCAs follows identifiable mechanisms, emerges under recognizable triggering conditions, and produces distinct failure modes. Elemental forgetting, concentration range collapse, phase-space amnesia, and temperature-induced overwrite are not isolated anomalies. They are predictable outcomes of parameter interference, representation drift, output-distribution shift, and gradient conflict.

The methodological consequence is clear. Sequential alloy learning must be evaluated with forgetting-sensitive metrics, benchmarked under controlled task sequences, and supported by mitigation strategies from the outset. Robust lifelong materials modeling requires knowledge retention to be treated as a design objective equal in importance to predictive accuracy. Under that shift, sequential learning can move from being a fragile approximation of continual discovery to a credible foundation for scalable modeling in high-entropy and multi-principal-element alloys.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 04 Jul 2021

Revised: 28 Sep 2021

Accepted: 03 Dec 2021

Published online: 18 January 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Rao Z, Tung PY, Xie R, Wei Y, Zhang H, Ferrari A, et al. Machine learning-enabled high-entropy alloy discovery. *Science*. 2022;378(6615):78-85.
- Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater*. 2019;170:109-17.
- Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, Aykol M, et al. Benchmarking the acceleration of materials discovery by sequential learning. *Chem Sci*. 2020;11(10):2696-706.
- McCloskey M, Cohen NJ. Catastrophic interference in connectionist networks: The sequential learning problem. In: *Psychol Learn Motiv*. 1989;24:109-65.
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, et al. Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci U S A*. 2017;114(13):3521-6.
- De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, et al. A continual learning survey: Defying forgetting in classification tasks. *IEEE Trans Pattern Anal Mach Intell*. 2022;44(7):3366-85.
- Hadsell R, Rao D, Rusu AA, Pascanu R. Embracing change: Continual learning in deep neural networks. *Trends Cogn Sci*. 2020;24(12):1028-40.
- Bartók AP, De S, Poelking C, Bernstein N, Kermode JR, Csányi G, et al. Machine learning unifies the modeling of materials and molecules. *Sci Adv*. 2017;3(12):e1701816.
- Zuo Y, Chen C, Li X, Deng Z, Chen Y, Behler J, et al. Performance and cost assessment of machine learning interatomic potentials. *J Phys Chem A*. 2020;124(4):731-45.
- Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun*. 2022;13(1):2453.
- Eckhoff M, Reiher M. Lifelong machine learning potentials. *J Chem Theory Comput*. 2023;19(12):3509-25.
- Chen S, Cheng Y, Gao H. Machine learning for high-entropy alloys. In: *Artificial intelligence for materials science*. Cham: Springer International Publishing; 2021. p. 21-58.
- Chen L, Chen Z, Yao X, Su B, Chen W, Pang X, et al. High-entropy alloy catalysts: High-throughput and machine learning-driven design. *J Mater Inform*. 2022;2(4):N-A.
- Li Y, Holmedal B, Liu B, Li H, Zhuang L, Zhang J, et al. Towards high-throughput microstructure simulation in compositionally complex alloys via machine learning. *Calphad*. 2021;72:102231.
- Murty BS, Yeh JW, Ranganathan S, Bhattacharjee PP. High-entropy alloys. Amsterdam: Elsevier; 2019.
- Müller MN, Brix C, Bak S, Liu C, Johnson TT. The third international verification of neural networks competition (VNN-COMP 2022): Summary and results. *arXiv preprint arXiv:2212.10376*. 2022 Dec 20.

Mirzadeh SI, Chaudhry A, Yin D, Hu H, Pascanu R, Gorur D, et al. Wide neural networks forget less catastrophically. In: International conference on machine learning; 2022. p. 15699-717.

Huang W, Martin P, Zhuang HL. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* 2019;169:225-36.

Zhang Y, Wen C, Wang C, Antonov S, Xue D, Bai Y, et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* 2020;185:528-39.

Wu Q, Wang Z, Hu X, Zheng T, Yang Z, He F, et al. Uncovering the eutectics design by machine learning in the Al-Co-Cr-Fe-Ni high entropy system. *Acta Mater.* 2020;182:278-86.

Liu Y, Parisot S, Slabaugh G, Jia X, Leonardis A, Tuytelaars T. More classifiers, less forgetting: A generic multi-classifier paradigm for incremental learning. In: European conference on computer vision. Cham: Springer International Publishing; 2020. p. 699-716.

Aljundi R, Rohrbach M, Tuytelaars T. Selfless sequential learning. *arXiv preprint arXiv:1806.05421*. 2018 Jun 14.

Giles SA, Sengupta D, Broderick SR, Rajan K. Machine-learning-based intelligent framework for discovering refractory high-entropy alloys with improved high-temperature yield strength. *npj Comput Mater.* 2022;8(1):235.

Kaufmann K, Vecchio KS. Searching for high entropy alloys: A machine learning approach. *Acta Mater.* 2020;198:178-222.