Improvement of CHF Prediction under Non-Uniform Axial Power Profiles through AI-Based Memory-Effect Correction

Hyunjun Jang, and Kyung Mo Kim*
Department of Energy Engineering, Korea Institute of Energy Technology (KENTECH),
21 Kentech-gil, Naju-si, Jeonnam 58330, Republic of Korea
hjjang1360@kentech.ac.kr, kmokim@kentech.ac.kr

***Keywords :** critical heat flux, non-uniform axial power, differentiable physics, model-informed neural network

1. Introduction

Critical heat flux limits the power of water-cooled reactors, and the practically relevant case is non-uniform axial heating, because real cores never operate with a flat power profile. Data-driven CHF models have achieved useful accuracy, but almost all of them share one structural choice. The CHF magnitude and the CHF location are predicted as two independent regression outputs. This choice discards the physics that connects the two quantities. In the experiment the power is raised slowly, the local wall heat flux profile grows, the local tolerance of the boiling channel shrinks as the coolant heats up, and the crisis occurs at the first contact between the two curves. The location is not a second unknown. It is the position of that contact.

The classical engineering answer to non-uniform heating is the correction-factor methodology. Tong expressed the upstream thermal memory as an exponential kernel with an empirical coefficient C [2], and later work by Hwang et al. derived the same functional form from bubble-layer conservation and showed that C should be a function of the local state rather than constant [3]. The look-up table of Groeneveld provides the uniform local limit that the correction acts on [1]. These elements are exactly the ingredients of the design methodology used in reactor safety analysis, where the minimum DNBR is searched along the channel.

A recent study in this laboratory [4] combined a look-up-table-informed deep ensemble with the F-factor methodology and reported a location R^2 of 0.99. One point requires care in reading that number. The reported configuration tunes correction parameters for each test case against the measured CHF and the measured location, with the answers inside the tuning objective. Such a value is the theoretical ceiling of parameterization, that is, the accuracy the framework can reach when the answer is available. It is not a prediction, because a new operating condition has no answer. The distinction matters, and it defines the goal of the present work: to approach that ceiling without looking at the answers.

This paper reports on three contributions. First, the ceiling itself is reproduced independently. When the same per-case calibration protocol is applied to a faithful reproduction of the reference solver, the absolute-location reaches 0.992, in agreement with the reported 0.99. This confirms that the ceiling is a property of the methodology and fixes the target. Second, four answer-free improvements are introduced and validated on a strictly held-out test set: (a) pre-state features, in which a preliminary simulation with no correction supplies four physically meaningful descriptors to the network that predicts the correction coefficients, (b) end-to-end training through the power search itself, implemented with an implicit-function gradient so that the training signal and the inference mechanism share the same physics, (c) a learned multiplicative correction g(x) of the limit-curve shape, and (d) a global positional correction shared by all cases. Third, one diagnostic principle organizes the study: a trade-off between the CHF error and the location error is read as evidence of model misspecification, because both quantities are read from the same margin curve. Applied repeatedly, the principle motivated the two global corrections and finally uncovered a numerical defect of the discretized F-factor whose repair improved every stage of the model, including the answer-using ceiling. The chain from metric diagnosis to learned compensation to root-cause repair is proposed as a transferable procedure for physics-informed predictors.

2. Methods

2.1. Data and conventions

Two databases are used. The uniform-heating database is the 2006 Groeneveld look-up-table compilation (24,579 tube tests) [1]. The non-uniform database is a KAERI compilation of 888 tests with fitted 50-node axial power profiles, of which 852 tests carry a recorded CHF location and form the prediction table. One documentation detail is essential and was verified on all 888 tests: the recorded CHF of the non-uniform database is the average heat flux at the critical condition, CHF = W/(DL), where W is the critical channel power. The critical power is therefore the natural output of a simulation, and the identity is satisfied by construction in everything that follows. The recorded locations lie on the discrete thermocouple grid of each test section, with a spacing of about 0.1 m for the 7 m series and the last station upstream of the outlet, which bounds the achievable location resolution. The table is split 70/15/15 (598/127/127) with a fixed seed. All learned components see only the training split, and every result below is reported on the held-out 127 tests, which include the most extreme CHF sample of the database and the full

range of heated lengths. Locations are reported in normalized form z/L, and in meters only where a comparison with absolute-coordinate literature values is made, because the absolute is inflated by the variance of the heated length. Ceiling (answer-using) values and prediction (answer-free) values are never mixed in one table without labels.

2.2. The physics simulator

For a candidate channel power, the solver evaluates the following quantities on the 50-node mesh of the fitted axial profile. The local equilibrium quality follows from the cumulative energy balance,

$$x(z) = \frac{h(z)-h_f(P)}{h_{fg}(P)}, h(z) = h_{in} + \frac{4}{GD}\int_0^z q''(z')dz' \quad (1)$$

where q"(z) is the local wall heat flux at the candidate power. The uniform local limit q_u(P, G, x(z), D) is evaluated from the 2006 look-up table with the standard diameter correction $\left(\frac{0.008}{D}\right)^{0.5}$ and a residual deep ensemble of ten networks trained on the uniform database. The upstream thermal memory follows the Tong formulation,

$$C(z) = 5.906\,\frac{(1-x(z))^{4.31}}{(G/1356)^{0.478}} \quad (2)$$

$$F(z) = \frac{C\int_{z_{ONB}}^{z} q''(z')e^{-C(z-z')}dz'}{q''(z)\left[1-e^{-C(z-z_{ONB})}\right]} \quad (3)$$

anchored at the onset of significant void and clipped to [0.5, 5]. The memory integral of Eq. (3) is discretized on the mesh with trapezoidal weights. Section 3.3 shows that a rectangle rule in its place creates a spurious spike of F at the onset point and that the spike dominated one third of the location predictions before it was identified. The effective correction combines a per-case positional multiplier with the two learned global corrections introduced in Sections 2.6 and 2.7,

$$F_{eff}(z) = F(z)e^{c_0\xi + c_1\xi^2}\varphi(\xi), \xi = z/L \quad (4)$$

where c_0 and c_1 are per-case coefficients and ϕ is global. The margin profile is

$$\mathrm{DNBR}(z) = \frac{q_u(P,G,x(z),D)g(x(z))}{F_{eff}(z)q''(z)} \quad (5)$$

The power is raised by bisection until the minimum of the DNBR profile reaches unity. The power at first contact W_c gives the CHF through the database convention,

$$\mathrm{CHF} = \frac{W_c}{\pi DL} \quad (6)$$

and the contact position gives the location, refined parabolically around the discrete minimum. Both outputs come from one calculation, exactly as in the mDNBR search for design practice.

2.3. The ceiling and the deployable gap

In the ceiling protocol the per-case parameters are fitted to each test against its own measured CHF and location. This uses the answer at inference time and is reported here only as the upper bound of the parameterization. In the deployable protocol a student network predicts the correction coefficients from inputs alone. The gap between the two protocols is the quantity this work reduces.

2.4. Answer-free improvement 1: pre-state features

The student receives thirteen inputs: seven operating descriptors, two shape descriptors, and four pre-state features computed by one preliminary simulation with no correction (c = 0): the uncorrected critical flux, the uncorrected contact position, the exit quality at that power, and the flatness of the margin curve. These quantities are computable without any measurement of the case. The rationale is that the amount of correction a case needs depends on which physical regime the case sits in, and the cheapest description of that regime is the model's own uncorrected answer.

2.5. Answer-free improvement 2: training through the solver

Distillation of per-case oracle labels loses information, because many coefficient values produce nearly the same fit and the labels are therefore noisy. Instead, the student is trained end to end through the power search. The critical power for the current network output is found by bisection and made differentiable with respect to the network parameters through an implicit-function correction,

$$\hat{q} = q^* - \frac{m(q^*)}{\partial m/\partial q|_{q^*}}, m(q) = \min_z \mathrm{DNBR}(z;q) - 1 \quad (7)$$

where the minimum is smoothed by a log-sum-exp with temperature 0.02, the location is read with a soft argmin with temperature 0.05, and the correction term is clipped to a trust region of 20% of q* for stability. The training objective is

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{q}_i - q_{exp,i}}{q_{exp,i}}\right)^2 + w_{loc}\frac{1}{N}\sum_{i=1}^{N}\left(l_{soft,i} - l_{exp,i}\right)^2 \quad (8)$$

so that the training signal and the inference mechanism share the same physics. Answers enter only the training loss. The second positional coefficient c1, unlearnable

from oracle labels, becomes usable through the task loss. End-to-end training from scratch fails, and the working recipe is distillation pre-training followed by solver-in-the-loop fine-tuning.

2.6. Answer-free improvement 3: the limit-shape correction g(x)

If the model physics were exact, the CHF and the location would improve or degrade together, because both are read from the same margin curve. A trade-off between them under re-weighting of the loss is therefore evidence that the curve shape is wrong, and the diagnosis points at the quality dependence of the local limit. A global multiplicative correction g(x) = exp(s(x)), with s a small network initialized at zero, is applied to the limit and trained jointly with the student through the solver. The falsifiable prediction is that, if the shape is the true deficiency, both outputs improve simultaneously with no location weighting.

2.7. Answer-free improvement 4: the global positional correction $\phi(\xi)$

The per-case multiplier of Eq. (4) lets the student adapt the axial correction to each case, but any bias shared by all cases, such as the treatment of the onset point or a mesh effect, is wasteful for the student to relearn case by case. A global multiplicative factor $\phi(\xi)$ = exp(s(ξ)), with s a small network initialized at zero, is therefore added on top of the per-case multiplier and trained jointly with the student and g(x) through the solver. The same falsifiable criterion applies. A consistent learnability pattern is observed for both global functions: when trained alone against a frozen student neither g(x) nor $\phi(\xi)$ moves away from unity, and the gain appears only under joint optimization, because the global component and the per-case component are entangled.

3. Results

3.1. Reproduction of the ceiling

With per-case calibration of a single coefficient against the answers, the faithful reproduction of the reference solver reaches an absolute-location of 0.992 and a location MAE of 0.106 m on the test set. The reported 0.9897 and 0.079 m of the five-parameter configuration in Ref. [4] are therefore confirmed as the methodology ceiling. The corresponding uncorrected baseline (0.814, 0.51 m) matches the look-up-table row of the same reference. One might conclude from the ceiling that the location problem is solved. It is not: without the answers the same parameterization falls to the deployable values below, and the remainder of this section concerns that regime only.

3.2. The answer-free ladder

Table I and Fig. 1 summarize the deployable accuracy on the held-out 127 tests under the corrected solver of Section 3.3. Figure 2 shows the parity of the final model, and Fig. 3 illustrates the contact mechanism at the predicted critical power for four heater shapes.

Table I: The answer-free ladder on the held-out test set. The last row is the answer-using oracle bound.

Configuration	**CHF MAPE (%)**	**Location MAE (z/L)**
Distilled student (7 inputs)	7.22	0.134
+ pre-state features (13 inputs)	7.92	0.115
+ end-to-end training through the solver	5.88	0.106
+ limit-shape correction g(x)	5.88	0.100
+ global positional correction $\phi(\xi)$	5.84	0.101
Ceiling (single-coefficient oracle)	5.00	0.088

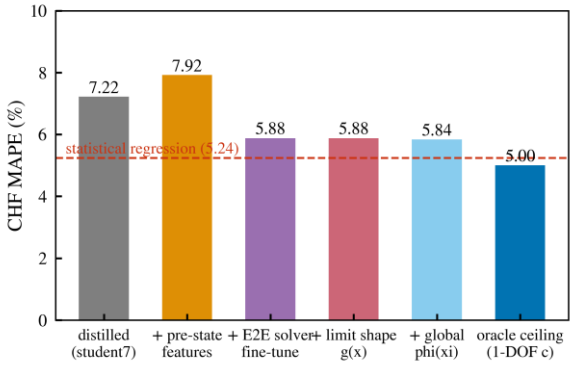


Fig. 1. Answer-free performance ladder on the held-out test set. The dashed line marks statistical regression reference.

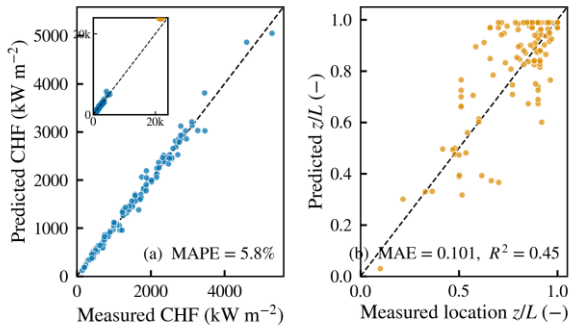


Fig. 2. Parity of the final model on the held-out test set: CHF (left, with the extreme sample in the inset) and normalized location (right).

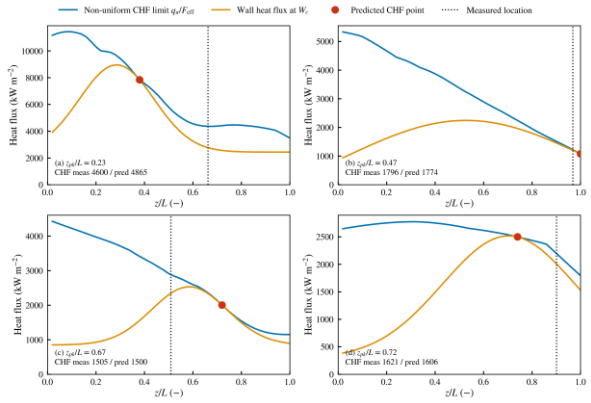


Fig. 3. Contact mechanism at the predicted critical power for four heater shapes. The crisis is predicted at the first contact between the wall heat flux and the corrected local limit. In (b) the power peak sits at mid-channel while both the measured and the predicted crisis lie near the outlet, because the location is set by the margin and not by the flux peak.

Three observations follow. First, the pre-state features act on the location axis: the preliminary-simulation descriptors reduce the location error from 0.134 to 0.115 at a small CHF cost, and Section 3.3 explains why the large CHF gain they showed under the defective solver has disappeared. Second, the largest single gain comes from end-to-end training through the solver, 7.92% to 5.88% in CHF error, which confirms that the training signal and the inference mechanism should share the same physics. Third, the global corrections contribute small but consistent refinements of the location error, 0.106 to 0.100, and the final model reaches 5.84% with a location error of 0.101 and a normalized-location R^2 of 0.45, within 0.84 percentage points and 0.013 of the single-coefficient oracle ceiling. In absolute coordinates the final location error is 0.373 m with R^2 = 0.897, the axis on which the literature values of Section 3.1 are quoted. The normalized-location R^2 should be read against the parameterization itself: the answer-using oracle reaches 0.58 on the same axis, and the residual scatter reflects the location information the framework does not carry.

3.3. The trade-off diagnosis and the quadrature defect

The corrected solver of Table I is itself a product of the diagnostic principle of Section 2.6, and this subsection reports that chain of evidence. Under the initial rectangle-rule solver the model showed a clear trade-off: raising the location weight in the loss improved the location error from 0.167 to 0.127 while degrading the CHF error from 8.4% to 11.2%. Read as misspecification rather than as a tuning opportunity, the trade-off motivated the two global corrections, and both produced the predicted signature of simultaneous improvement, 8.36% to 8.01% to 7.19% in CHF error together with 0.167 to 0.136 to 0.120 in location error.

The same principle then demanded inspection of the margin curve itself, and the curve showed a V-shaped collapse of the corrected limit at the first mesh node past the onset point. The correction factor jumped from 1.000 to 2.003 in one node and relaxed as (n+1)/n over the following nodes. This is the exact signature of endpoint double counting: a rectangle-rule discretization of Eq. (3) assigns a full cell width to both end nodes of the integration interval, so that the numerator covers two cells while the denominator spans one. Physically F must approach unity continuously at the onset, because no bubble-layer history has accumulated there. The spurious spike halved the effective limit at a single node and captured the minimum of the margin curve: 34% of the test predictions placed the crisis exactly at that node, and those cases carried a 43% larger location error than the rest. A smooth global function such as $\phi(\xi)$ cannot represent the defect, because the spike stands at a different position in every case.

Replacing the rectangle rule with trapezoidal endpoint weights removed the spike, and every stage of the model improved when retrained (Table II). The ceiling itself rose from 6.74% to 5.00% in CHF error and from 0.118 to 0.088 in location error, which shows that the defect had been capping the parameterization rather than the learning. The fraction of predictions trapped at the onset node fell from 34% to zero. After the correction the contributions of g(x) and $\phi(\xi)$ shrink to a fraction of their former size, which identifies what the learned corrections had been doing: absorbing the smooth average component of the defect. One might conclude that the learned corrections were therefore superfluous. They were not. They were the instruments that localized the defect, and the sequence of trade-off detection, learned compensation, and root-cause repair is proposed as the transferable methodological result of this work.

Table II: Effect of the quadrature correction on every stage of the model (CHF MAPE / location MAE). Both columns are measured on the same held-out test set.

Stage	**Rectangle rule (% / z/L)**	**Trapezoidal rule (% / z/L)**
Ceiling (oracle)	6.74 / 0.118	5.00 / 0.088
Distilled student (13 inputs)	9.47 / 0.160	7.92 / 0.115
+ end-to-end training	8.36 / 0.167	5.88 / 0.106
+ g(x)	8.01 / 0.136	5.88 / 0.100
+ $\phi(\xi)$	7.19 / 0.120	5.84 / 0.101

3.4. Position of statistical reference

A purely statistical ensemble (a five-member heteroscedastic deep ensemble trained on the same split,

predicting CHF and location as two regression outputs) reaches 5.24% with a location MAE of 0.045, and a stacked variant that receives the four pre-state physics features reaches 4.71%. The physics-simulation model now stands within 0.6 percentage points of this in-distribution reference, which can absorb facility-level systematic effects the physics model is structurally forbidden from memorizing, while one third of the test cases have a near neighbor in the training split. The advantages of the simulation model lie on axes the regression does not possess: the location follows from the physics instead of a second regression head, arbitrary axial profiles are accepted, extrapolation follows the physical structure, and the learned quantities remain interpretable as a modernization of the Tong coefficient and a closure correction. Synthetic augmentation by a physics-informed diffusion model [5], which does repair the sparse high-CHF tail of the statistical predictor, was also evaluated and added nothing to the physics-simulation predictor at any ratio, a null result identified as saturation of the low-dimensional correction.

4. Discussion

The central methodological point is the use of a metric trade-off as a misspecification detector. A weighting dial that moves accuracy between CHF and location is not a tuning opportunity, it is a measurement that the margin curve is wrong. In this work the detector fired three times at increasing depth. It first motivated the quality-axis correction g(x), then the position-axis correction $\phi(\xi)$, and finally forced a direct inspection of the margin curve that uncovered a discretization defect no smooth global function could represent. Acting on the root cause was worth more than all learned compensations combined, and the residual roles of the learned components after the repair, location refinement for the pre-state features and for the global corrections, measure how much genuine closure error remains.

The remaining distance to the oracle ceiling is 0.84 percentage points in CHF error and 0.013 in normalized location. This residual is information the thirteen inputs do not carry, and the ceiling itself, 5.00% and 0.088, bounds what any coefficient predictor can achieve inside the present parameterization. Part of the location residual is not model error at all: the measured locations lie on a thermocouple grid of about 0.1 m whose last station sits upstream of the outlet, while the simulator reports a continuous coordinate.

Four limitations are stated plainly. The results are single-seed and single-split. The test set, while strictly held out, is not a new facility, and one third of its cases have near neighbors in training because the underlying experiments vary the heater shape at matched conditions. The corrections g(x) and phi(xi) are entangled with the student re-adaptation, so their shapes should be read as learned closure adjustments rather than as measured properties of the limit curve and the axial closusre. Above a local quality of about 0.9 the limit rests on the dryout cliff of the look-up table and on residual-ensemble extrapolation, a region containing three percent of the test contacts, so its shape carries no physical meaning. The reproduction values of Section 3.1 were obtained with the pre-correction solver.

5. Conclusions

A CHF prediction model was built that reproduces the physical process of the measurement: the power is raised until the wall heat flux first touches the corrected local limit, and the CHF and its location are obtained from that contact in a single calculation. The accuracy ceiling of the underlying F-factor parameterization was established by reproducing the per-case calibrated results of prior work (location 0.99), and that ceiling was then approached without using answers at inference. Four learned components carry the model: pre-state features from a preliminary simulation, end-to-end training through a differentiable power search, and two global closure corrections. One diagnostic principle carries the methodology: a trade-off between the CHF error and the location error is evidence of misspecification. The principle led from loss re-weighting experiments to the learned global corrections and finally to the discovery and repair of an endpoint double-counting defect in the discretized Tong integral, which had been capping even the answer-using ceiling. On a strictly held-out test set that includes the most extreme sample of the database the final model reaches a CHF error of 5.8% and a location error of 0.101 in z/L (0.37 m and R2 = 0.897 in absolute coordinates), within 0.6 percentage points of a purely statistical ensemble, with the location produced by the physics of the power ramp instead of a second regression head.

ACKNOWLEDGEMENT

This work was supported by the Innovative Small Modular Reactor Development Agency grant (No. RS-2023-00257680) and Nuclear Safety Research Program through the Korea Foundation of Nuclear Safety (KoFONS) funded by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No. RS-2026-25538149).

REFERENCES

[1] D. C. Groeneveld, Critical Heat Flux Data Used to Generate the 2006 Groeneveld Lookup Tables, NUREG/KM-0011, U.S. NRC, 2019.
[2] L. S. Tong, Prediction of Departure from Nucleate Boiling for an Axially Non-Uniform Heat Flux Distribution, Journal of Nuclear Energy, Vol. 21, pp. 241-248, 1967.
[3] D. H. Hwang et al., bubble-layer derivation of the non-uniform CHF correction factor, 2001. [full bibliographic details to be verified]
[4] C. Y. Yoon and K. M. Kim, Deep Ensemble Algorithm for CHF Prediction under Uniform/Nonuniform Axial Profile Conditions, Transactions of the Korean Nuclear Society Autumn Meeting, 2025.
[5] A. Akins and X. Wu, Physics-Informed Diffusion Model for Generation of Physically Consistent Critical Heat Flux Data, Transactions of the American Nuclear Society, 2026.