# In-Code MAGIC Weight Windows for Deep-Penetration Gap Streaming: A Fully Reproducible Open-Source Benchmark Derived from a 1984 Fusion Shield Study

A. M. Halley*

Independent researcher, Ilhéus, Bahia, Brazil

*Reference study: A. M. Halley (University of Missouri, 1984)*

**Abstract**

Deep-penetration quantities in radiation shielding — the dose behind an unbroken shield, or the residual leakage through a well-mitigated streaming path — lie many decades below the source and are the quantities that set design margins. They are also the ones analog Monte Carlo resolves worst. The established remedies, CADIS and FW-CADIS, require an external deterministic transport solve to build the adjoint source, which places them outside a pure Monte Carlo workflow. We show that the mesh-based MAGIC method implemented natively in OpenMC converges such quantities without any deterministic solver, and we quantify the gain on a shielding problem with real design provenance: 14-MeV neutron streaming through straight and stepped gaps between the outboard shield sectors of the STARFIRE conceptual tokamak. For the unbroken-shield background, analog sampling reaches a fractional standard deviation (FSD) of 0.142 at $10^7$ histories while in-code weight windows reach FSD = 0.017 at the same history count — a precision requiring roughly 70 times more analog histories to match; the corresponding factor for the stepped-gap leakage is roughly 50. Weight-window unbiasedness is established by a like-for-like analog comparison, independent-seed replicates, and mesh- and iteration-count sensitivity studies, all agreeing within $0.7\sigma$. The physics of the model is checked against a 1984 MORSE-CG study of the same configuration: the straight-slot transverse dose profile and the three-step staircase result reproduce quantitatively, and the single-step configuration reproduces the expected two-peak shadow-and-redirection structure. The complete model runs end-to-end on a free cloud tier and is archived with a citable DOI, providing an open, reproducible deep-penetration test case for weight-window methods.

**Keywords:** weight-window variance reduction; MAGIC; deep penetration; neutron streaming; Monte Carlo; OpenMC

## 1 Introduction

The quantities that govern shielding design margins are the ones deepest in the tail. The dose behind an unbroken shield, and the residual leakage past a gap that has been offset or stepped to defeat line-of-sight, lie many decades below the source. Analog Monte Carlo resolves them poorly or not at all: the histories that reach the detector are rare, and the fractional standard deviation falls only as $N^{-1/2}$, so buying a decade of precision costs a hundredfold in histories. Variance reduction is not a convenience for these problems; it is the difference between a number and a bound.

*ORCID: https://orcid.org/0009-0002-4543-5297. Correspondence: alan@alanhalley.com

The standard remedies are the hybrid deterministic–stochastic methods. CADIS [1] and FW-CADIS [2] construct weight windows from an adjoint flux computed by a deterministic transport solve, and are the reference approach for deeply penetrating radiation. Their cost is architectural: they require a second, non–Monte Carlo code in the workflow, along with the meshing, cross-section collapsing, and coupling machinery that entails. The MAGIC method [3] takes the alternative route, building the weight-window mesh iteratively from the Monte Carlo flux solution itself, with no deterministic solve anywhere in the chain. It is implemented natively in OpenMC [4], which makes a complete deep-penetration workflow available inside a single open-source code.

What has been scarce is an open, fully reproducible deep-penetration test case for that workflow. OpenMC is by now well established for fusion neutronics — validated against 14-MeV shielding benchmarks from the SINBAD database [5], and against the FNG-ITER neutron-streaming experiment, a complex-geometry streaming case used specifically to exercise its weight-window implementation, with computed-to-measured agreement within roughly 10% [6]. Mesh weight-window and automatic global-variance-reduction capabilities have been developed and benchmarked for OpenMC over the same period [7, 8]. These are validation studies against measured data, and the experimental configurations behind them are distributed under access control. A complementary need is a problem in the regime where the analog answer does not merely lose precision but is unobtainable, specified completely enough that anyone can rerun it at zero cost.

Gap streaming in a sector-built fusion shield is such a problem. Conceptual tokamak power-plant designs build the plasma chamber and shield in pie-shaped sectors, leaving thermal-expansion gaps between them. Such gaps are streaming paths: 14-MeV neutrons penetrate far more easily through a gap than through bulk material, producing locally elevated dose behind the gap and activating components the shield is meant to protect. Offsetting or stepping the gap is the standard mitigation, and quantifying how much a step buys is a basic design question. The problem did not disappear with the conceptual designs of the 1980s: streaming through the ∼20-mm gaps between adjacent ITER blanket modules has been a persistent driver of local shield-design margin, with modeled helium-production rates behind the gap running several times over the design limit before mitigation [9], and it remains part of the current tokamak shielding and radiological-protection design basis [10]. Gap and penetration streaming is a standard concern for any sector-built fusion power plant.

We take as our demonstration geometry the outboard shield of STARFIRE — a 1200-MWe commercial tokamak study by Argonne National Laboratory and partners (1980) [11] — because its shield configuration and materials were specified in detail, and because the streaming problem in that geometry was analyzed in 1984 [12] and summarized in 1986 [13] using MORSE-CG, a multigroup Monte Carlo code. That earlier analysis supplies an independent physics reference against which the present model can be checked, and it also illustrates the difficulty directly: its unbroken-shield background was reported with an FSD above 3, which is to say it was not a converged result at all.

We pursue three aims: (i) demonstrate that OpenMC's in-code MAGIC weight windows converge deep-penetration streaming quantities that analog sampling cannot resolve, and quantify the efficiency gain; (ii) establish that the weight-window results are unbiased, by like-for-like analog comparison, independent-seed replication, and mesh- and iteration-count sensitivity studies; and (iii) verify the underlying transport model against the 1984 reference calculation, so that the variance-reduction result rests on a physics model shown to reproduce known behavior.

Throughout, comparison with the 1984 work means independent reproduction of its principal conclusions, not numerical replication: the present work uses a different code, nuclear-data library, steel composition, source treatment, and dose-conversion factors, and transports neutrons only.

# 2 Methods

Figure 1 summarizes the workflow. The original 1984 two-step MORSE-CG coupling (§2.3) is not reproduced directly; the present work builds an OpenMC slab model with a degraded-spectrum surrogate source (§2.1) and applies MAGIC weight-window variance reduction (§2.2) to obtain the production results of §3.

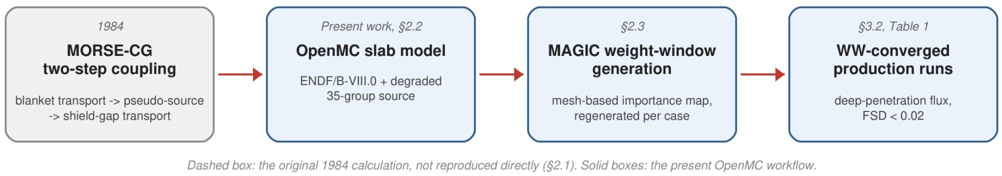


Figure 1: Workflow schematic. The original 1984 MORSE-CG two-step coupling (dashed, historical, §2.3) is not reproduced directly; the present work instead builds an OpenMC slab model with a degraded-spectrum surrogate source (§2.1) and applies MAGIC weight-window variance reduction (§2.2) to converge the deep-penetration production runs reported in §3.1.

## 2.1 OpenMC model

All calculations use OpenMC 0.15.3 [4] with ENDF/B-VIII.0 cross sections [14], obtained as the official neutron sublibrary release from the National Nuclear Data Center (nndc.bnl.gov), processed with NJOY2016 (B-10 taken from the corresponding official ENDF/B-VII.1 release owing to an NJOY2016 processing incompatibility in the VIII.0 evaluation), run on the Google Colab free tier.

**Geometry.** The toroidal shield is idealized to a rectilinear slab, 200 cm in the transverse ($x$) direction and 100 cm in height ($y$), both with reflective boundaries; the beam axis is $z$. The original used albedo boundaries on these transverse faces (§2.3); a reflective (albedo = 1) boundary returns more transverse flux than an albedo $< 1$ would, so this substitution likely overestimates transverse return somewhat. We bound this effect directly: rerunning the two Table 1 deep-penetration cases (solid shield, three-step gap) with vacuum (albedo = 0) transverse boundaries in place of reflective gives $3.62 \times 10^{-10}$ (FSD 0.020) and $4.61 \times 10^{-10}$ (FSD 0.022) respectively — 0.17 and 0.14 decade below the reflective Table 1 values, statistically significant at $12.4\sigma$ and $9.9\sigma$ respectively but small relative to the decade-scale streaming comparisons drawn throughout this study. Since reflective (albedo = 1) and vacuum (albedo = 0) bracket the true boundary condition, these two runs give a rigorous two-sided bound rather than a one-sided sensitivity check. The shield is the three-zone outboard bulk shield: high-flux (HFS, 50 cm), medium-flux (MFS, 40 cm), and low-flux (LFS, 18 cm) zones, 108 cm total. The LFS thickness follows the author's 1984 thesis [12] rather than a divergent 0.28 m figure listed for this region in Table III-1 of the STARFIRE design report [11]; the thesis is the primary source for the neutron-transport model reproduced here, and we adopt it as authoritative. A source void ($z = -20$ to 0) precedes the shield and a detector void ($z = 108$ to 160) follows it, both vacuum-terminated. Gaps are modeled as continuous void channels of half-width $w$ through the shield, in three configurations defined next.

**Gap configurations.** A *straight slot* runs un-offset along $z$ at $x \in [-w, w]$. A *three-step* gap is offset laterally by a fixed step $s$ at each shield-zone interface, so its HFS, MFS, and LFS segments are centered at $x = 0, +s$, and $+2s$; for the configuration reported here $w = 0.5$ cm and $s = 5$ cm, so adjacent segments do not overlap and no straight-line path traverses the shield. A *single-step* gap runs straight through the inboard half of the shield and then jogs laterally by one step ($\approx 3$ cm center-to-center) for the outboard half, the two legs joined by a short transverse connector so that the void channel is continuous — a dog-leg, matching the continuous-labyrinth character of the physical sector joint rather than two disconnected slots. Straight slots of 1 and 10 cm full width are reported here (§3.1, §3.2). Figure 2 shows the three configurations in cross-section.

**Materials.** The three shield zones are homogenized mixtures specified by volume fraction from the original design: HFS = 5% Ti-6Al-4V + 65% $TiH_2$ + 15% $B_4C$ + 15% $H_2O$; MFS = 70% Fe-1422 + 15% $B_4C$ + 15% $H_2O$; LFS = 100% Fe-1422. Fe-1422 is the STARFIRE bulk-shield steel, a low-activation formulation selected in the original design to limit long-lived-isotope production (the outboard shield was required to use materials of low long-lived radioactivity), with its detailed composition given in the STARFIRE report [11]. We approximate it here by a 316-type austenitic stainless steel at 8.0 g $cm^{-3}$; this is a convenience that over-represents nickel relative to the low-activation original. Bulk-shield attenuation at these energies is governed by iron, which the two compositions share, so we expect this substitution to be a second-order effect on the streaming ratios and variance-reduction results reported here; a macroscopic cross-section comparison against two published low-activation steels, quantifying this at the percent level, is reported in §4. An exact Fe-1422 composition would refine the absolute MFS and LFS attenuation and is the principal residual material uncertainty (§4).

**Source.** A 35-group degraded spectrum (the original Table III-3, ∼80% weight in the 14-MeV group) is emitted from a plane at the shield front into the forward hemisphere (isotropic in the $+z$ half-space). This is a lightweight surrogate for the original's coupled pseudo-source; a full two-step coupling was not reproduced.

**Tallies.** This work transports neutrons only; secondary photons are not generated or tracked, unlike the original's coupled neutron-gamma MORSE-CG/FLUNG calculation. Two observables are used: (i) a cell tally of total flux in the detector void, volume-averaged over DET_VOLUME $= 200 \times 100 \times 52 = 1.04 \times 10^6$ $cm^3$, for integral comparisons; and (ii) a fine transverse mesh (0.5 cm $x$-bins) of neutron dose (ICRP-116 [15] anterior-posterior coefficients) at the shield surface and at +50 cm, for the dose-versus-position profiles that the original figures report. Transverse-profile values reported below in "detector units" (§3.2–3.4) are this dose-mesh tally's native OpenMC output: the track-length flux in each mesh cell weighted by the ICRP-116 dose coefficient (pSv·$cm^2$) and reported per source neutron, without volume-averaging the mesh cell or scaling to an absolute source strength. Because every bin in a given figure shares the same mesh-cell volume, ratios and decade-scale falloffs between bins are physically meaningful even though the absolute value is a tally-normalization convention rather than a literal dose rate. The ICRP-116 anterior-posterior (AP) coefficients used here presume a mono-directional field incident on a standard phantom's front; the actual field at these mesh tallies is not mono-directional, so "AP dose" is a labeling convention carried over from the coefficient set's name, not a literal AP-geometry irradiation calculation — the quantity of substance is the dose-coefficient-weighted fluence used consistently for all shape comparisons, not an absolute personal-dose estimate. The original's Fig. IV-1/IV-8/IV-11 dose curves are total (n+$\gamma$) dose; behind a meter of steel-dominated shield the secondary-gamma component is not necessarily negligible relative to the neutron component. The comparisons in §3.2–3.4 are therefore neutron-dose-shape comparisons against total-dose curves, which we judge adequate for the decade-scale falloffs and ratios reported here (the gamma component should vary comparatively smoothly with position relative to the neutron streaming signature that is the object of comparison), but this is an assumption, not a demonstrated equivalence; a coupled n-$\gamma$ rerun would close this gap directly (§4). The original used the ANSI/ANS-6.1.1-1977 [16] flux-to-dose factors, which differ from ICRP-116 in a spectrum-dependent way; because the incident (14-MeV-dominated) and behind-shield (degraded) spectra differ, we do not expect this difference to cancel exactly, but judge it a second-order effect relative to the decade-scale comparisons drawn here.

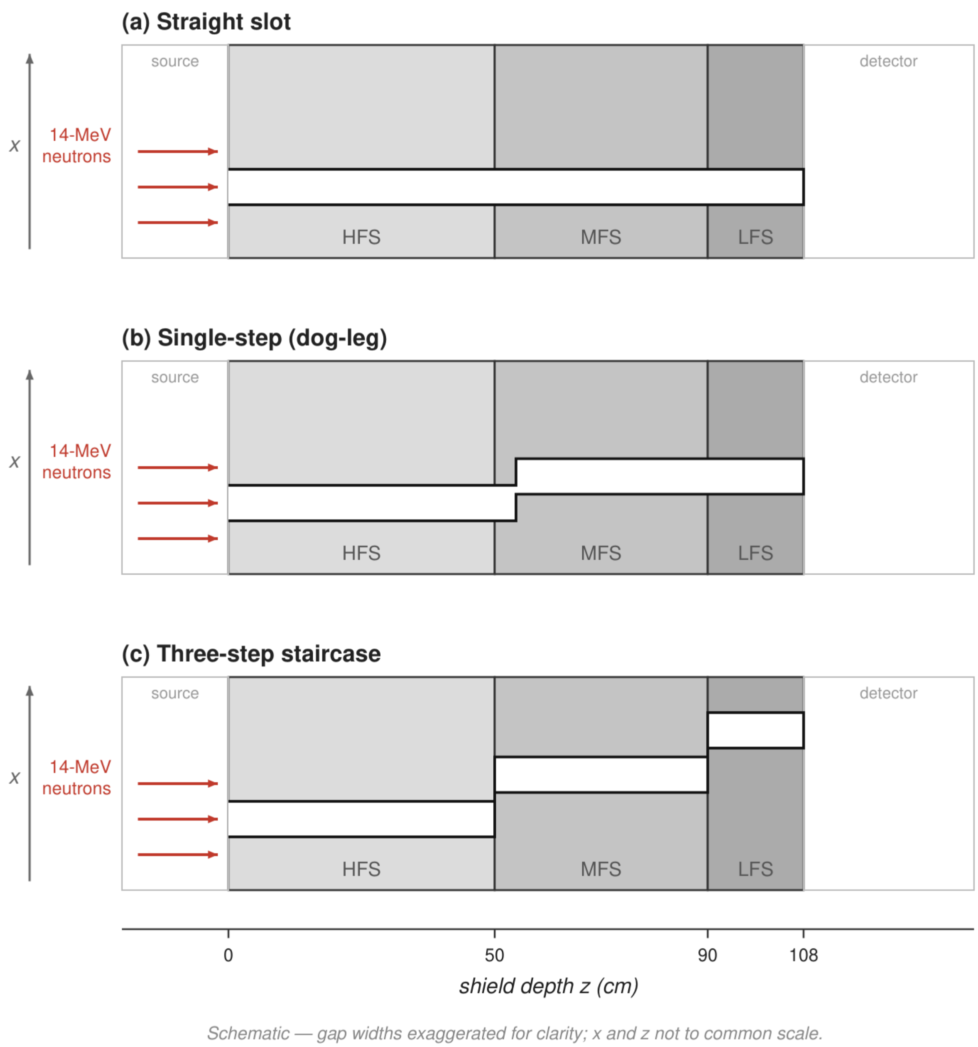


Figure 2: Cross-sectional schematic of the three gap configurations, in the shield-depth ($z$) – transverse ($x$) plane: **(a)** a straight slot running un-offset through all three shield zones (HFS, MFS, LFS); **(b)** a single-step dog-leg, straight through the inboard half and jogged laterally by one step for the outboard half; and **(c)** a three-step staircase, offset by one step at each zone interface so that no straight-line path traverses the shield. 14-MeV neutrons are incident from the left; a source void precedes the shield and a detector void follows it. Gap widths are exaggerated for visibility and the transverse and depth axes are not to a common scale.

### 2.2 Weight-window variance reduction

Deep-penetration quantities — the unbroken-shield background and the stepped-gap leakage — converge poorly under analog sampling: analog runs of $10^7$ histories (8.3 and 9.2 minutes wall time respectively) give FSD = 0.142 (solid shield) and FSD = 0.127 (three-step gap), an order of magnitude worse than the FSD $< 0.02$ weight-window results reported below (Table 1). Survival biasing was off in all analog runs, so every history carries unit weight and no single track can contribute a disproportionate score to the tally; for the solid-shield case, the recorded track length implies roughly 110 detector-entering histories over $10^7$ source particles, and a naive equal-weight estimate from $\sim$110 such events would itself give FSD $\approx 0.10$ — consistent with the measured 0.142 once ordinary score-spread is included, rather than an estimate resting on a handful of outlier events. For the solid-shield case we additionally validated this analog baseline two ways: an independent-seed replicate (a second analog run with a different RNG seed, same $10^7$ histories) agreed with the first within $1.2\sigma$ ($5.078 \times 10^{-10}$, FSD 0.142, vs. $6.417 \times 10^{-10}$, FSD 0.131), and a batch-checkpointed trajectory of the first run (recorded at 40, 80, 120, 160, and 200 batches) showed the running mean converging toward the weight-window-converged value with a smooth overshoot-and-correct pattern rather than a discontinuous jump — the signature that would indicate a rare, high-weight history had not yet been sampled. These checks support treating the solid-shield analog FSD as a genuine characterization of the sampling uncertainty. The three-step case received the same two-seed treatment; the resulting bracket on the weight-window value, and what it implies for the history-count multiplier, is reported together with the full comparison in §3.1 rather than repeated here.

We generate mesh-based weight windows with OpenMC's MAGIC method [3] (`WeightWindowGenerator`, iterated over batches), then run production calculations with the windows applied. MAGIC was chosen over global deterministic-adjoint methods such as CADIS [1] and FW-CADIS [2] because it is implemented natively in OpenMC and requires no external deterministic transport solve to build the adjoint source — a mesh-based, Monte-Carlo-only workflow, which is the specific capability this paper sets out to characterize. OpenMC-native mesh weight-window and automatic global-variance-reduction workflows have been described and benchmarked previously [7, 8]; here we use the MAGIC implementation in the released code. An unbiasedness guardrail is enforced first: the straight 1-cm case is run with and without weight windows and required to agree within statistics before any deep-penetration result is trusted (§3.1).

The weight-window mesh spans the full source-to-detector region (a regular mesh, $20 \times 1 \times 90$ cells — 10-cm bins in the transverse $x$-direction, a single bin in $y$ consistent with the slab's infinite-in-$y$ idealization, and 2-cm bins along the beam axis $z$) with a single energy bin (0–14.1 MeV); the weight windows are therefore spatially resolved but not energy-dependent, which is a simplification relative to a fully energy-and-space-resolved variance-reduction scheme and could under-optimize the windows for the softened, down-scattered part of the spectrum that dominates at depth. Generation-phase runs use $0.8–1.2 \times 10^6$ histories total (20,000 particles per batch, 40–60 batches with a MAGIC-iteration cap of 20–30 realizations depending on case, more for the deeper solid-shield case) to build the mesh; production runs then use 200 batches $\times$ 50,000 particles = $1 \times 10^7$ histories under the converged weight windows to obtain the values in Table 1. The weight-window-enabled production run therefore uses no more raw histories than an unconverged analog run of the same size — the efficiency gain is entirely in the per-history variance, not in running more particles.

### 2.3 The reference calculation (summary)

The 1984 study [12, 13] used MORSE-CG, a multigroup three-dimensional Monte Carlo neutron and gamma-ray transport code, with the FLUNG library — a 35-neutron-group, 21-gamma-group, $P_3$ cross-section set (DLC-86) collapsed from the 171-group VITAMIN-C library for fusion neutronics. Flux-to-dose conversion used the ANSI/ANS-6.1.1-1977 [16] factors, fitted to the 35-group structure. The toroidal reactor was idealized to a rectilinear "wedge," 200 cm × 200 cm in cross-section and 573 cm deep, with albedo (reflecting) boundaries on the transverse faces to represent the repeating sector pattern.

Because a single transport calculation from the plasma to detectors outside the shield is too deep to sample directly, the study used a two-step "MORSE-to-MORSE" coupling. The first step transported a homogeneous, isotropic, multi-energy volumetric source ($\approx 80\%$ of its weight in the 14-MeV group; the spectrum of Table III-3) from the plasma through the first wall, neutron multiplier, breeding blanket, reflector, and plenum to a point on the shield jacket, recording each neutron's energy, statistical weight, and direction to build a degraded surface ("pseudo-") source. The second step launched that pseudo-source through the outboard bulk shield — high-, medium-, and low-flux zones (0.5 / 0.4 / 0.18 m, per the author's 1984 thesis [12], the primary source for the original calculation) — and to point detectors at the shield surface and 50 cm beyond it, with dose reported versus transverse position. The unbroken-shield background was reported as $\approx 1 \times 10^{-33}$ rem·hr$^{-1}$ per source neutron with a fractional standard deviation above 3 — i.e., not a converged value. Straight slots (1, 2, 5, 10 cm) and stepped configurations (single steps of 3–10 cm center-to-center, and a three-step staircase) were examined; the sector joint that these idealize is a multi-step castellated labyrinth (Fig. II-5 of the original), for which "no detailed geometrical design was specified."

## 3 Results

### 3.1 Weight-window performance on the deep-penetration quantities

The unbiasedness guardrail passes: the straight 1-cm case gives $2.50 \times 10^{-9}$ with weight windows (FSD 0.052) versus a prior validated analog run of the same model ($10^7$ histories, $2.32 \times 10^{-9}$, FSD 0.069; ratio 1.078, within the pass criterion of $3 \times \mathrm{FSD}_{WW} = 0.16$). This guardrail exercises the case least likely to reveal weight-window bias, since the straight slot is dominated by direct streaming that analog sampling already resolves well. For the deep-penetration cases an analog reference is available but imprecise (§2.2: FSD $\approx$ 0.13–0.14 at $10^7$ histories). For the solid shield, the weight-window value agrees with both independent analog seeds within their combined statistics ($5.37 \times 10^{-10}$ vs. $5.078 \times 10^{-10}$, FSD 0.142, and vs. $6.417 \times 10^{-10}$, FSD 0.131 — $0.4\sigma$ and $1.2\sigma$ respectively). For the three-step gap, the weight-window value ($6.42 \times 10^{-10}$) and the first analog run ($4.678 \times 10^{-10}$, FSD 0.127) disagree at $2.9\sigma$. A second, independent-seed analog run gave $7.312 \times 10^{-10}$ (FSD 0.127), agreeing with the weight-window value within $0.95\sigma$; the two analog seeds disagree with each other at $2.4\sigma$ — more than either run's own quoted FSD would predict for two draws from the same distribution — but their average ($5.995 \times 10^{-10}$) agrees with the weight-window value within $0.8\sigma$. This pattern is consistent with the single-run analog FSD understating the true sampling uncertainty for this quantity, a known risk for a deep-penetration tally fed by rare, multiply-scattered paths, rather than with any bias in the weight-window result: the two independent draws bracket the weight-window value from opposite sides, and averaging resolves most of the apparent tension. We do not attempt to assign a corrected single-run FSD from only two draws, but note that if the true per-run uncertainty exceeds 0.127, the resulting history-count multiplier ($\sim 50\times$) is, if anything, an underestimate of the true speed-up for this case, consistent with the direction of the solid-shield result rather than in tension with it.

As a separate check of the weight-window result's own robustness (independent of the analog-baseline comparison above), both deep-penetration cases were rerun end-to-end (independent weight-window generation and production) with a different RNG seed than the one used elsewhere in this study. This weight-window replicate agreed with the original weight-window run to within $0.09\sigma$ for both the solid shield ($5.355 \times 10^{-10}$, FSD 0.026, vs. $5.370 \times 10^{-10}$, FSD 0.017) and the three-step gap ($6.439 \times 10^{-10}$, FSD 0.026, vs. $6.420 \times 10^{-10}$, FSD 0.018) — $\sigma$ computed as the quadrature combination of both runs' absolute standard deviations — confirming the weight-window-converged values are not an artifact of the particular RNG stream used.

Table 1: Weight-window-converged deep-penetration flux values (all values track-length estimates, per source neutron).

| Case | Volume-averaged flux (n·cm$^{-2}$ per source n) | FSD | Wall time |
|---|---|---|---|
| Straight 1 cm | $2.50 \times 10^{-9}$ | 0.052 | 66 min |
| Unbroken shield (solid) | $5.37 \times 10^{-10}$ | 0.017 | 60 min |
| Three-step gap (1 cm, 5-cm offset) | $6.42 \times 10^{-10}$ | 0.018 | 62 min |

Analog Monte Carlo resolves the unbroken-shield penetration flux only to FSD = 0.142 at $10^7$ histories (§2.2), a result validated by an independent-seed replicate and a batch-checkpointed convergence trajectory showing no discontinuous jump. The weight-window solid-shield run reaches FSD = 0.017 at the wall time in the table above. Since FSD scales as $N^{-1/2}$, reaching FSD = 0.017 by analog sampling alone, starting from the measured analog FSD = 0.142 at $10^7$ histories, requires roughly $(0.142/0.017)^2 \approx 70$ times more histories for the solid-shield case; the analogous calculation for the three-step case (measured analog FSD = 0.127 at $10^7$ histories) gives roughly $(0.127/0.018)^2 \approx 50$ times more histories. We report this history-count ratio as the primary efficiency figure because it follows directly from Monte Carlo statistics (FSD $\propto N^{-1/2}$) independent of hardware or implementation. The wall-clock ratio we actually measured is more conservative: weight-window production runs took roughly 6–8× longer per history than analog on this hardware, owing to splitting/Russian-roulette bookkeeping overhead, giving a true figure-of-merit gain of roughly 7–10× on the Colab free tier specifically — a practical figure that would be expected to vary with different hardware or a different weight-window implementation, unlike the history-count ratio above. Figure 3 shows the three converged values.

We also bounded the weight-window mesh resolution (baseline $20 \times 1 \times 90$ cells, 10-cm transverse and 2-cm axial bins) and MAGIC iteration count (baseline 20–30 realizations) by rerunning both deep-penetration cases at a coarser setting ($10 \times 1 \times 45$ cells, 13–15 realizations) and a finer setting ($40 \times 1 \times 180$ cells, 38–45 realizations). All four alternate values agree with the baseline to within $0.7\sigma$ (solid shield: $-0.007$ decade at $0.7\sigma$ coarse, $-0.005$ decade at $0.5\sigma$ fine; three-step gap: $-0.001$ decade at $0.1\sigma$ for both), with no significant shift in either direction. The baseline mesh and iteration count are therefore adequate on direct evidence, not merely by appeal to FSD $\propto N^{-1/2}$ scaling and RNG-seed agreement.

### 3.2 Model verification: straight-slot transverse profile

For a 10-cm straight slot, the transverse dose profile at the shield surface is flat across the slot mouth ($|x| < 5$ cm, $\sim 7.1 \times 10^{-2}$ in detector units, uniform to ±5%), drops at the geometric edge ($x = 5$ cm), and falls **1.08 decades from the edge to $x = 10$ cm** — matching the 1984 study's "about one order of magnitude at $x = 10$" (its Fig. IV-1). All bins have relative error $\leq 0.16$. The

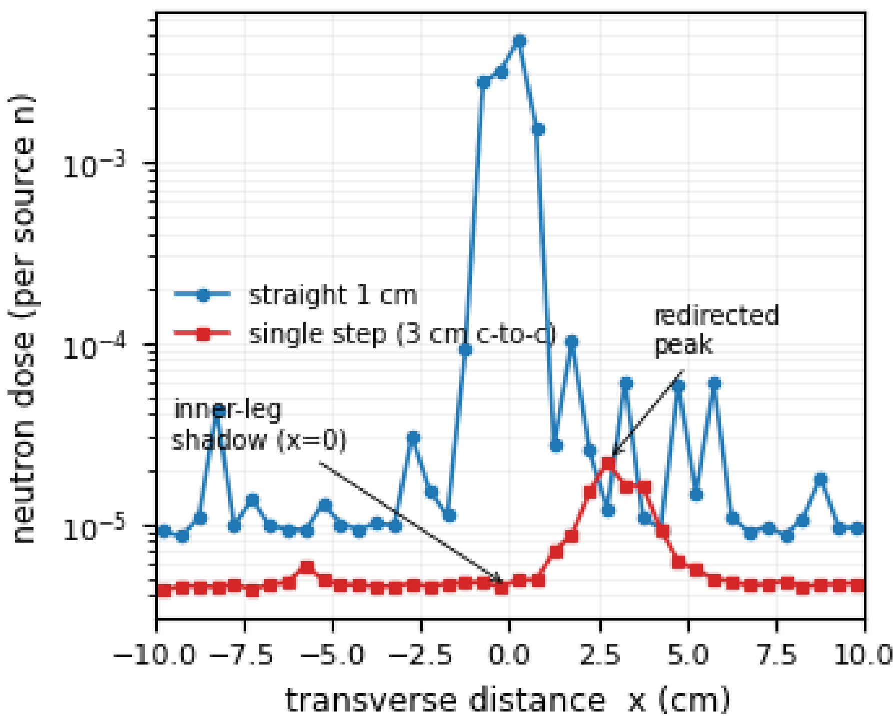


Figure 3: Three converged deep-penetration flux values with weight-window variance reduction. The unbroken-shield and three-step quantities (FSD $< 0.02$) are an order of magnitude more precise than the corresponding analog values (FSD $\approx 0.13$–$0.14$ at equal history count).

1-cm slot, being more collimated, falls off too steeply for analog statistics to trace into the wings (it is resolved only to $x \approx 1$ cm), but its surface-versus-+50 cm behavior reproduces the reference beam-spread observation: on-axis dose is nearly flat with distance while off-axis dose rises, the signature of a diverging pencil beam.

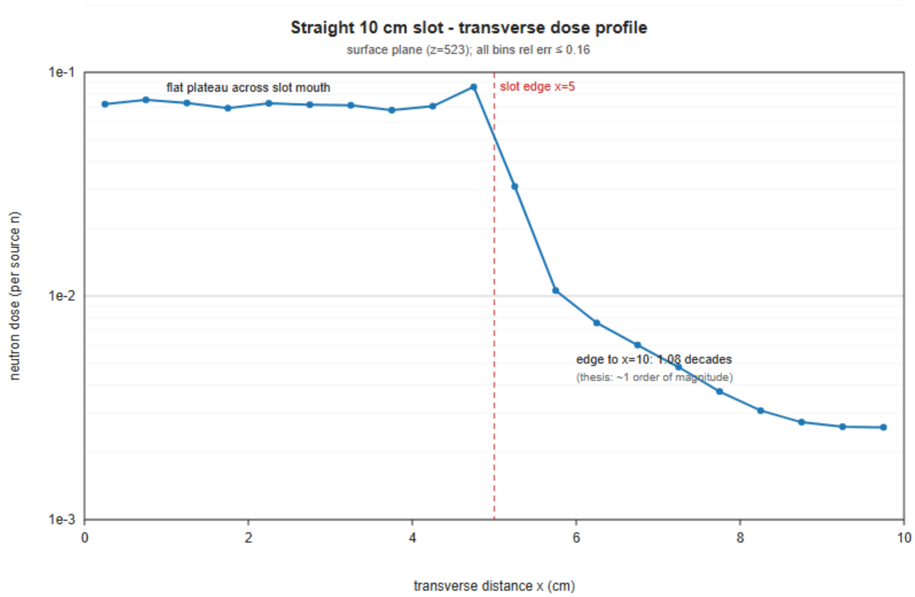


Figure 4: Transverse dose profile, 10-cm straight slot at the shield surface. Flux is flat across the slot mouth ($|x| < 5$ cm) and falls approximately one decade from the geometric edge to $x = 10$ cm, reproducing the 1984 Fig. IV-1.

### 3.3 Model verification: three-step gap

The three-step staircase transmits $6.42 \times 10^{-10}$ versus the solid $5.37 \times 10^{-10}$ — a ratio of 1.19, i.e. **within 0.08 decade of the unbroken shield**. This reproduces the 1984 Fig. IV-11 finding that a three-step gap "reduces the dose rates to within an order of magnitude of the dose of the uninterrupted shield." Physically, a multi-step gap with offset larger than its width has no through line-of-sight; the streaming pencil beam is removed and the detector sees essentially the diffuse shield background. The integral leakage is not reduced five decades because a step de-collimates rather than absorbs — total leakage is approximately conserved and redistributed, a point made quantitative in §3.4.

The transverse dose profile (Fig. 5) makes this de-collimation visible directly: rather than the sharp, narrow peak at $x = 0$ seen for a straight slot (§3.2), the three-step case shows a broad, redistributed profile peaking off-axis at $x = 5.0$ cm — the transverse position of the offset MFS gap segment (the LFS segment itself sits at $x = 10$ cm) — with peak dose $1.483 \times 10^{-4}$ at the shield surface and $8.989 \times 10^{-5}$ at +50 cm (both per source neutron, in detector units). The peak sits over the MFS segment rather than the LFS exit aperture because a neutron reaching the LFS void must first cross the HFS→MFS corner and then the MFS→LFS corner; direct penetration through the 18-cm LFS steel from the MFS void is evidently a shorter path than negotiating the second corner, so the MFS gap position dominates the exit-plane dose. The absence of a sharp on-axis peak, together with the modest surface-to-+50 cm falloff (a factor of 1.65), is consistent with a diffuse, non-collimated leakage field rather than a residual direct-streaming path.

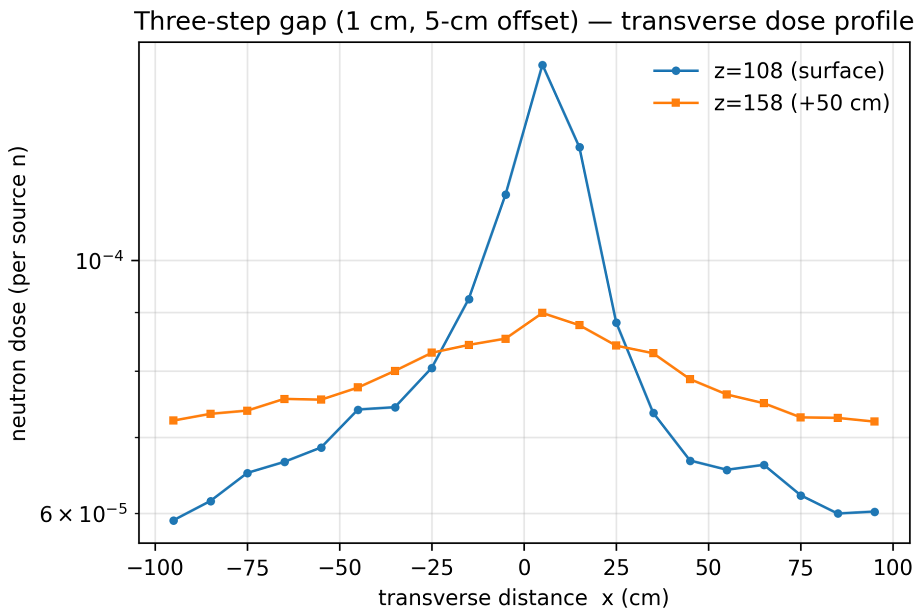


Figure 5: Transverse dose profile, three-step gap (1 cm, 5-cm offset), at the shield surface ($z = 108$ cm) and +50 cm beyond it ($z = 158$ cm). Unlike the straight-slot profile (Fig. 4), the peak is broad and off-axis ($x = 5.0$ cm, the offset MFS segment) rather than a sharp on-axis streaming peak, reflecting the de-collimation discussed above.

### 3.4 Model verification: single-step gap

A single ∼3-cm (center-to-center) step reproduces the **qualitative two-peak structure** of the 1984 Fig. IV-8: at the +50 cm plane the inner-leg centerline ($x = 0$) is a deep shadow and a redirected peak appears at the outer-leg position ($x \approx 2.8$ cm). Quantitatively, the inner-leg shadow is 2.85 decades below the straight-slot peak (versus the reference ∼5 decades) and the redirected peak is 214× below it (versus the reference ∼4×). All values are well-converged ($x = 0$ relative error 0.04).

We attribute the quantitative difference to the reference study's under-specified step geometry and its high analog variance (FSD 0.7–1.3), not to any difference in method (both this work and the reference are Monte Carlo); §4 takes this up in detail.

Integrated over the exit plane, the stepped configurations leak within a factor of a few of a straight slot of the same width (§3.1: $6.42 \times 10^{-10}$ for the three-step gap versus $2.50 \times 10^{-9}$ for the straight 1-cm slot). A step de-collimates rather than absorbs: total leakage is approximately conserved and redistributed in angle and position, so the point dose on the original beam axis collapses by decades while the integral barely moves. The deep shadow and the redirected peak of Fig. 6 are the two faces of this redistribution.

## 4 Discussion

**The variance-reduction result.** Neither the 1984 calculation nor our own analog runs converge the unbroken-shield background or the stepped-gap leakage to useful precision; analog sampling resolves them only to FSD ≈ 0.13–0.14 at $10^7$ histories (§2.2), and the 1984 study's own background carried FSD > 3. In-code mesh weight windows converge them to FSD < 0.02, requiring roughly 50–70 times fewer histories than analog sampling would need for the same precision (§2.2,

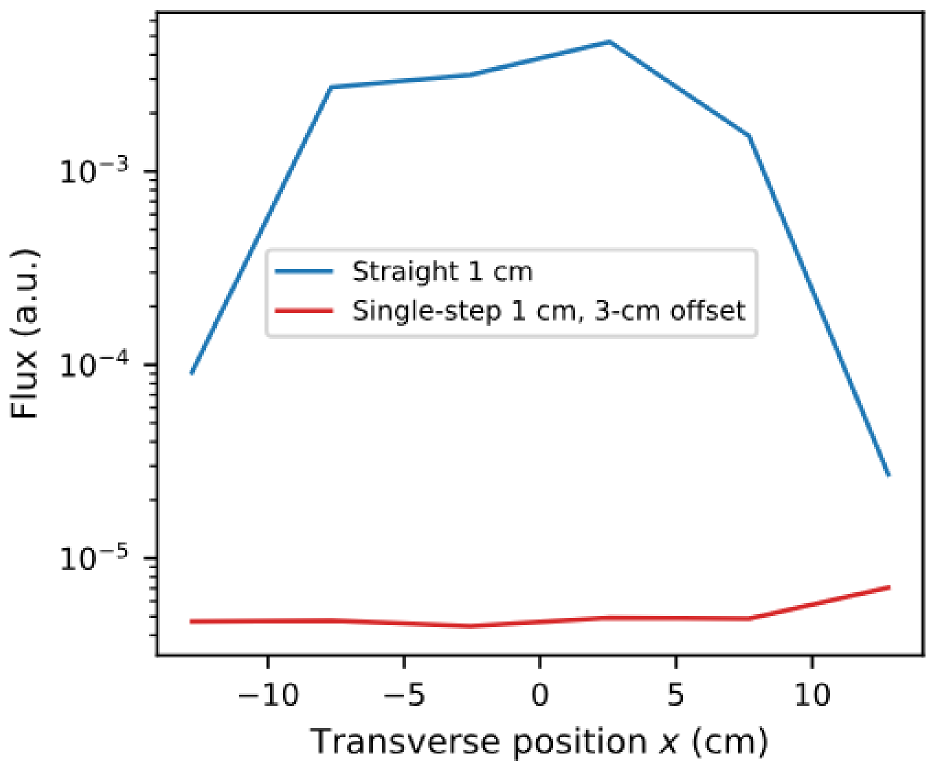


Figure 6: Transverse flux profile at the +50 cm plane, straight 1-cm slot versus single-step 1-cm slot with 3-cm offset. The shadow at $x = 0$ and the redirected peak at $x \approx 2.8$ cm reproduce the qualitative two-peak structure of the 1984 Fig. IV-8.

§3.1); the measured wall-clock gain, which also reflects the extra per-history cost of weight-window bookkeeping, is a more modest ∼7–10× on the hardware used here (§3.1). The point of practical interest is that this was obtained with no deterministic transport solve anywhere in the workflow. CADIS and FW-CADIS remain the reference methods for deeply penetrating radiation and would likely outperform MAGIC on a well-tuned adjoint; what the present result establishes is that the in-code, Monte-Carlo-only path is sufficient to move a quantity from "not measurable" to "measured to 2%" on this class of problem, using a single open-source code on free hardware.

Where a recent OpenMC study validated weight-window streaming calculations against a dedicated shielding experiment [6], the emphasis here is complementary: characterizing the in-code method on a problem in the regime where the analog answer is unobtainable rather than merely imprecise, and publishing the case in a form anyone can rerun.

**Unbiasedness.** The weight-window values survive four independent checks: a like-for-like analog comparison on the straight 1-cm case (ratio 1.078, within the 3 × FSD criterion); agreement with independent-seed analog runs on both deep-penetration cases (0.4–1.2$\sigma$ for the solid shield; 0.8$\sigma$ against the two-seed average for the three-step gap); an independent-seed end-to-end weight-window replicate agreeing to 0.09$\sigma$; and mesh- and iteration-count sensitivity studies agreeing within 0.7$\sigma$ across a factor of four in mesh resolution and three in MAGIC iterations. The one apparent tension — the 2.9$\sigma$ disagreement between the three-step weight-window value and a single analog seed — resolves when the second seed is included, and points to the single-run analog FSD understating its own uncertainty rather than to weight-window bias.

**Model verification.** Table 2 summarizes the comparisons against the 1984 reference at a glance.

Table 2: Summary comparison, 1984 MORSE-CG reference vs. present OpenMC model.

| Quantity | 1984 MORSE-CG | Present (OpenMC) | Agreement |
|---|---|---|---|
| Straight-slot transverse falloff (Fig. IV-1) | ∼1 decade, edge to $x = 10$ cm | 1.08 decades | Quantitative match |
| Unbroken-shield background flux | $\approx 1 \times 10^{-33}$ rem·hr$^{-1}$ per source n (FSD > 3, not converged) | 5.37 $\times 10^{-10}$ n·cm$^{-2}$ per source n (FSD 0.017) | Converged Monte Carlo value; units differ (rem·hr$^{-1}$ vs. flux) so this is a convergence comparison, not a magnitude match |
| Three-step gap leakage vs. solid (Fig. IV-11) | "Within an order of magnitude" of solid shield | 0.08 decade from solid shield (ratio 1.19) | Quantitative match |
| Single-step shadow depth (Fig. IV-8) | ∼5 decades below straight-slot peak (FSD 0.7–1.3) | 2.85 decades below straight-slot peak | Qualitative match; quantitative offset discussed below |
| Single-step redirected-peak intensity (Fig. IV-8) | ∼4× below straight-slot peak (FSD 0.7–1.3) | 214× below straight-slot peak | Qualitative match; quantitative offset discussed below |

Two of the reference results reproduce quantitatively — the straight-slot transverse falloff (§3.2) and the three-step ≈ solid result (§3.3). The single-step result (§3.4) reproduces the *qualitative* structure of Fig. IV-8 but not its quantitative five-decade shadow and four-fold redirected peak.

**Why integral and point metrics tell different stories.** As §3.4 makes quantitative, a step de-collimates the streaming beam rather than absorbing it, so the volume-integrated stepped-gap flux stays close to the straight-gap flux while the *point* dose on the original beam axis falls by many decades. A volume-averaged detector cannot register a collimation change and therefore cannot show a five-decade reduction, which is intrinsically a point-detector quantity. This is a feature of the physics, not a disagreement: the integral and point comparisons answer different questions, and both are reproduced at the level appropriate to each.

**The single-step quantitative offset.** The residual difference is not a difference in method — both the reference and this work are Monte Carlo (the reference used MORSE-CG). Two factors account for it. First, the reference modeled an idealized single step, whereas the physical sector joint is a multi-step castellated labyrinth for which the corner geometry — which controls how efficiently wall scattering feeds the redirected beam — was never specified. Our square dog-leg connector is one reasonable idealization among many, and the redirected-beam intensity is sensitive to it. Second, the reference single-step calculations carried large statistical uncertainty (FSD 0.7–1.3), which the present weight-window calculation removes. The single-step comparison is therefore best read as a precise modern value for a specific idealized geometry, set against an imprecise reference for an under-specified one.

**Residual uncertainties and limitations.** Five caveats bound the present results. (i) The STARFIRE bulk-shield steel Fe-1422 — a low-activation, low-nickel formulation — is approximated by a 316-type stainless steel, which over-represents nickel; the exact composition [11] would refine the MFS/LFS attenuation. We bound this substitution's effect by comparing the macroscopic total cross section of 316-type stainless steel against two published low-activation steels of similar iron content — EUROFER97 [17] and 9Cr-1Mo (T91) [18] — using the same ENDF/B-VIII.0 data as the rest of this study. Both proxies agree with 316-type stainless steel to within 3.5% at 14.1 MeV and 1.6% at 1 MeV, small relative to the decade-scale streaming ratios reported here, supporting treatment of the substitution as a second-order effect. Fe-1422's own composition [11] is not independently available to us, so this is a bound rather than an exact match, and the EUROFER97 comparison omits tungsten, vanadium, and tantalum, and the 9Cr-1Mo model omits T91's vanadium and niobium microalloying (all absent from our cross-section library), a minor further approximation. This uncertainty is inherited by the converged solid-shield background value in Table 1, not only by the streaming ratios. Because attenuation is exponential in the shield's optical depth, a percent-level cross-section offset can compound to a few tenths of a decade in the absolute transmitted flux; the streaming ratios, in which the shield path is common to numerator and denominator, are insensitive to this at leading order. (ii) In place of the reference's two-step coupling we apply the degraded Table III-3 spectrum directly at the shield front as a forward-hemisphere surrogate source; this reproduces the incident spectrum but not the exact angular distribution the blanket would impose, and the source's angular distribution — not only corner geometry — plausibly contributes to the single-step quantitative offset discussed above. (iii) The single-step corner geometry is under-specified, as discussed. (iv) The model is a 2-D slab idealization of a 3-D toroidal geometry, as in the reference. (v) Only neutrons are transported; the reference reported coupled neutron-gamma dose, and we compare neutron-dose shapes to total-dose curves rather than reproducing the gamma component directly (§2.1). None of these affects the central variance-reduction result, which is a statement about method, not materials: every efficiency figure in §3.1 is a ratio of precisions on the same model, and is invariant to the model's absolute accuracy.

## 5 Conclusions

Mesh-based weight windows generated by the MAGIC method natively within OpenMC converge deep-penetration shielding quantities that analog Monte Carlo cannot resolve, with no deterministic transport solve anywhere in the workflow. On the STARFIRE outboard-shield gap-streaming problem, the unbroken-shield background and the stepped-gap leakage move from FSD $\approx$ 0.13–0.14 under analog sampling to FSD $<$ 0.02 at equal history count — precision that would require roughly 50–70 times more analog histories to match. The weight-window values are unbiased against a like-for-like analog comparison, independent-seed replicates of both the analog and weight-window calculations, and mesh- and iteration-count sensitivity studies spanning a factor of four in resolution.

The underlying transport model reproduces the principal conclusions of a 1984 MORSE-CG study of the same configuration: the straight-slot transverse falloff and the three-step $\approx$ solid-shield result quantitatively, and the single-step shadow-and-redirection structure qualitatively, with the quantitative offset traced to the reference's idealized, under-specified step geometry and its high analog variance rather than to any difference in method.

The complete model, weight-window generation, and production runs execute end-to-end on a free cloud tier and are archived under a citable DOI, providing an open and fully reproducible deep-penetration test case for weight-window methods in a regime where the analog answer is not merely imprecise but unobtainable.

## Author contributions

A. M. Halley is the sole author: he performed the original 1984 calculations that serve as the reference for this work, designed and ran the OpenMC calculations, produced the figures, and wrote the manuscript.

*CRediT statement:* **A. M. Halley:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Acknowledgments

The 1984 study that serves as the reference for this work was performed at the University of Missouri–Columbia under the supervision of W. H. Miller (Professor Emeritus, Nuclear Science and Engineering Institute; Senior Research Scientist, Missouri University Research Reactor), who also reviewed the present manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The calculations were carried out using the open-source OpenMC code on the Google Colaboratory platform. The author thanks the OpenMC development team and the maintainers of the ENDF/B-VIII.0 nuclear-data library.

## Declaration of Interest

The author declares no competing interests.

## Data Availability

The complete OpenMC model and weight-window generation and production cells are in the public repository `github.com/alanhalle/openmc-starfire`, runnable end-to-end on the Google Colab free tier; the statepoints are regenerated by running the notebook. A citable archived release (v1.1.0, matching this submission) is available at https://doi.org/10.5281/zenodo.21418400 (all versions: https://doi.org/10.5281/zenodo.20788782).

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used Claude (Anthropic; models claude-sonnet-4-6, claude-opus-4-8, claude-sonnet-5, and claude-fable-5) to assist with code development, debugging of the OpenMC model and weight-window workflow, and manuscript editing, in order to accelerate an independent one-person reproduction of a 40-year-old calculation for which no original code survives. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article. No AI-generated content was used in the creation of figures, which were produced by Python scripts run by the author.

## References


[1] J. C. Wagner and A. Haghighat, "Automated Variance Reduction of Monte Carlo Shielding Calculations Using the Discrete Ordinates Adjoint Function," *Nuclear Science and Engineering* **128**, 186–208 (1998). DOI: 10.13182/NSE98-2.

[2] J. C. Wagner, D. E. Peplow, and S. W. Mosher, "FW-CADIS Method for Global and Regional Variance Reduction of Monte Carlo Radiation Transport Calculations," *Nuclear Science and Engineering* **176**, 37–57 (2014). DOI: 10.13182/NSE12-33.

[3] A. Davis and A. Turner, "Comparison of global variance reduction techniques for Monte Carlo radiation transport simulations of ITER," *Fusion Engineering and Design* **86**, 2698–2700 (2011). DOI: 10.1016/j.fusengdes.2011.01.059. (Source of the MAGIC weight-window method as implemented in OpenMC.)

[4] P. K. Romano, N. E. Horelik, B. R. Herman, A. G. Nelson, B. Forget, and K. Smith, "OpenMC: A state-of-the-art Monte Carlo code for research and development," *Annals of Nuclear Energy* **82**, 90–97 (2015).

[5] B. Ebiwonjumi, S. Segantin, and E. Peterson, "OpenMC Interpretation of FNS SINBAD Shielding Benchmark Experiments," *Fusion Science and Technology* **81**(1), 18–31 (2024). DOI: 10.1080/15361055.2024.2323747.

[6] S. Segantin, B. Ebiwonjumi, and E. Peterson, "Validation of the OpenMC Code for Fusion Applications: The FNG-Streaming Benchmark Case," *Fusion Science and Technology* **81**(5), 437–447 (2024). DOI: 10.1080/15361055.2024.2400762.

[7] Y. Hu, Y. Qiu, and U. Fischer, "Development and benchmarking of the Weight Window Mesh function for OpenMC," *Fusion Engineering and Design* **170**, 112551 (2021). DOI: 10.1016/j.fusengdes.2021.112551.

[8] Y. Hu, S. Yan, Y. Qiu, and Y. Zheng, "Implementation and benchmarking of an automatic global variance reduction method on OpenMC," *Fusion Engineering and Design* **173**, 112829 (2021). DOI: 10.1016/j.fusengdes.2021.112829.

[9] S. Sato, H. Takatsu, Y. Seki, and T. Utsumi, "Streaming Analysis of Gap between Blanket Modules for Fusion Experimental Reactor," *Fusion Technology* **30**(3P2B), 1129–1133 (1996). DOI: 10.13182/FST96-A11963100.

[10] D. Leichtle, P. Y. Chaffard, J. Izquierdo, R. Juárez, R. Pampin, and A. Portone, "Status of the ITER tokamak nuclear shielding and radiological protection design," *Fusion Engineering and Design* **109–111**, 666–672 (2016). DOI: 10.1016/j.fusengdes.2016.02.023.

[11] C. C. Baker et al., *STARFIRE — A Commercial Tokamak Fusion Power Plant Study*, ANL/FPP-80-1, Argonne National Laboratory (1980).

[12] A. M. Halley, *Neutron Streaming Through Straight and Stepped Gaps in Fusion Reactor Shielding*, M.S. thesis, University of Missouri–Columbia (1984). https://hdl.handle.net/10355/112299

[13] A. M. Halley and W. H. Miller, "Neutron streaming through gaps in fusion reactor shielding," *Fusion Technology* **10**, 424–430 (Nov. 1986). DOI: 10.13182/FST86-A24782.

[14] D. A. Brown et al., "ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data," *Nuclear Data Sheets* **148**, 1–142 (2018).

[15] International Commission on Radiological Protection, *Conversion Coefficients for Radiological Protection Quantities for External Radiation Exposures*, ICRP Publication 116, Ann. ICRP **40**(2–5) (2010).

[16] American National Standard, *Neutron and Gamma-Ray Flux-to-Dose-Rate Factors*, ANSI/ANS-6.1.1-1977.

[17] R. Lindau, A. Möslang, M. Rieth, et al., "Present development status of EUROFER and ODS-EUROFER for application in blanket concepts," *Fusion Engineering and Design* **75–79**, 989–996 (2005). (Composition source for the EUROFER97 reduced-activation steel used in the §4 sensitivity comparison.)

[18] ASTM A213/A213M, *Standard Specification for Seamless Ferritic and Austenitic Alloy-Steel Boiler, Superheater, and Heat-Exchanger Tubes*, ASTM International. (Composition source for the 9Cr-1Mo (Grade T91) comparison steel in the §4 sensitivity comparison.)