Monash University
Department of Materials Science and Engineering

Dr Yuefeng Yin
AEA Ignite Project Lead / Research Fellow
yuefeng.yin@monash.edu

InterfaceHandler: Complex Interface for Semiconductors

NCMAS 2027 Application — Category 2 (Standard)

FACILITY
NCI: Gadi

CATEGORY
Category 2 (Standard)

CALENDAR YEAR
2027

RESOURCE REQUESTED
3,000 kSU (3 MSU) compute; 2 TiB /g/data; 1 TiB /scratch (default)

Part 1 — Research Proposal

Submitted separately as its own PDF; limited to 5 pages under NCMAS 2027 Category 2 rules.

1. Background and goals

Low-energy electronic devices — topological transistors, spin-orbit logic, and negative-capacitance field-effect transistors — depend on stacking dissimilar two-dimensional (2D) and thin-film materials into heterostructures whose interface physics (band alignment, orbital hybridisation, strain and disorder tolerance) determines whether a device concept survives contact with a real, manufacturable stack. My group's ongoing work (AEA Ignite Round 2, "From Lab to Fab"; ARC Linkage Project LP250200919, in partnership with TQ Transistors) has identified several promising material families — Heusler ferromagnetic thin films (Co_2MnGa-type), 2D bismuth allotropes, and topological pyrite-type crystals — but progress has been limited by a structural bottleneck: constructing, matching and screening the very large number of candidate heterostructure interfaces required to move from a single hero calculation to a validated design rule.

To remove this bottleneck we have built **InterfaceHandler (InHand)**, an open-source Python package that automates the generation, lattice-matched supercell construction, and electronic-structure-aware screening of heterostructure interfaces from first-principles-ready structural files. InHand is now feature-complete for Phase 1 (structure generation; current release 2.1.9) and is used in this proposal as the computational front end to a high-throughput DFT screening campaign on NCI's Gadi. The goal for the 2027 calendar year is to use InHand-generated heterostructure libraries to (i) screen several thousand candidate stacking combinations across three material families for type-I/type-II band alignment and disorder tolerance, (ii) validate the highest-priority candidates with converged plane-wave DFT, and (iii) generate the labelled dataset needed to train the machine-learning surrogate models that form Stream 3 of my broader research program.

1.1 The three material families and their current limits

Heusler ferromagnetic thin films (Co_2MnGa-type). We have shown that these films retain large Berry curvature even under complete structural disorder, and that compositional tuning ($Co_2MnGa_{1-x}Ge_x$) controls spin-wave propagation. What is not yet known is whether this robustness survives when the film is placed in contact with a realistic capping or substrate layer rather than modelled in isolation — the geometry every actual spintronic device requires.

2D bismuth allotropes. Five years of work in this family — from the original Wannier-function tight-binding models through to edge-state stabilisation in 2D topological crystalline insulators — has established the physics of an isolated bismuthene layer in detail. The open question for device relevance is edge- and interface-state survival against a realistic contact/substrate stack, which requires screening across many candidate contact chemistries rather than the single substrates modelled to date.

Topological pyrite-type crystals (OsX_2, X = Se, Te). Prior work established selective control of surface spin current in the free-standing crystal. Translating this into a usable spin-injection or spin-detection device again requires systematic

interface screening against candidate electrode/tunnel-barrier materials — precisely the combinatorial problem InHand is built to solve.

2. Significance, innovation and impact

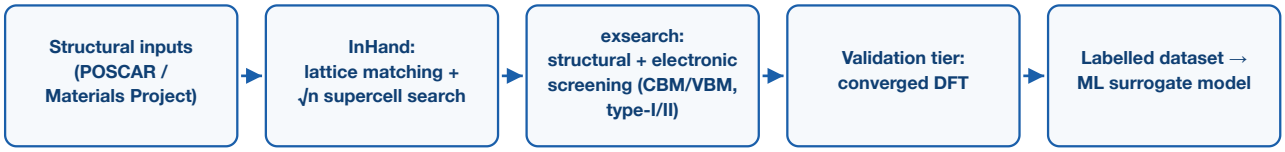

Figure 1. InHand's end-to-end pipeline: structural input, lattice-matched supercell generation, band-alignment-aware screening, DFT validation, and the labelled dataset that seeds the ML surrogate model (Section 3, Stage 4).

2.1 Innovation in the computational framework

Interface engineering in 2D and thin-film heterostructures is presently rate-limited by manual, one-structure-at-a-time workflows: a researcher hand-builds a POSCAR, checks the mismatch, and submits a single VASP job. This does not scale to the combinatorics of real materials discovery, where a three-layer ABA or ABC stack drawn from even a modest library of ten candidate 2D materials already implies hundreds of symmetry-inequivalent stacking sequences, each with several competing supercell solutions at different lattice-mismatch/atom-count trade-offs. InHand's `exsearch` module resolves this by encoding the selection problem explicitly: given a target stacking pattern (e.g. "XY/YX" heterobilayers, "ABA"/"ABC" trilayers with a tunnelling middle layer), it enumerates all symmetry-consistent supercell solutions (including √3, √7, √13 and √19 hexagonal supercells), ranks them by a physically motivated priority order — type-I band alignment first, then insulating middle-layer character, then band-gap-difference and lattice-mismatch/atom-count — and returns the best *N* structures with a full structural and electronic report. This is, to our knowledge, the first open, general-purpose tool that couples supercell lattice-matching directly to band-alignment-aware electronic screening criteria (CBM/VBM, type-I/II/III contact) rather than treating structure generation and electronic filtering as separate manual steps. The same framework also codifies common 2D building blocks (MX_2 polymorphs, graphene, A/B/C stacking sequences) and a reversed "structure-to-parts" mode that decomposes an arbitrary input POSCAR back into its constituent layers — a capability we use below to re-analyse existing experimental and literature structures at scale rather than only generating new ones.

2.2 Significance and impact in the scientific domain

The scientific payoff is a systematic, reproducible map of which interface chemistries in the Heusler-thin-film, bismuth-allotrope and topological-pyrite families sustain their functional electronic character (Berry curvature, protected edge states, spin texture) once real interfacial disorder, strain and band bending are accounted for — extending my group's existing single-system results into a general, transferable design rule rather than a set of isolated case studies. A design rule of this kind — "which contact chemistries preserve type-I confinement and topological character for family X, and by how much does that tolerance degrade with lattice mismatch" — is reusable well beyond the three families studied here, and the underlying InHand package is released openly so other groups working on 2D and thin-film heterostructures can apply the same screening methodology to their own material systems.

2.3 Significance and impact on society and industry partners

This work sits directly upstream of two active translation projects. Under the AEA Ignite award ("From Lab to Fab: Advancing a New Low-Energy Transistor Towards Large-scale Manufacturing", Lead CI, A$650,000) and the ARC Linkage Project with industry partner TQ Transistors (A$1.7M, Key Participant), the practical question is which interface stacks are worth fabricating and testing experimentally. Screening thousands of candidates computationally before committing cleanroom and characterisation time is the difference between an efficient translation pipeline and one that burns fabrication budget on structures whose band alignment was never going to work. A shortlist of validated, disorder-tolerant interface designs delivered in 2027 will feed directly into device prototyping already underway with TQ Transistors, supporting the shared goal of a manufacturable low-energy transistor architecture and the broader case for a domestically designed low-energy transistor technology.

2.4 Alignment with national research priorities

The project advances Australia's published priorities in advanced manufacturing and critical/enabling technologies: it directly supports a nationally funded (AEA) transistor-manufacturing translation pathway, builds sovereign open-source computational infrastructure (InHand) rather than relying on closed commercial tools, and generates the labelled electronic-structure dataset needed for AI-enabled materials discovery — an explicit national research priority — without which the machine-learning surrogate-modelling component of this program could not proceed.

3. Proposed scope of work for the 2027 calendar year

The 2027 program follows the same four-stage pipeline across all three material families, executed in parallel rather than sequentially so that early results from one family inform screening priorities in the others. Family-specific detail is given first, followed by the quarterly schedule.

- **Heusler thin films:** contact layers drawn from candidate non-magnetic metals and wide-gap insulating capping layers relevant to spin-injection device geometries; primary screening criterion is preservation of Berry curvature and spin-wave character across the interface.
- **2D bismuth allotropes:** contact layers drawn from hexagonal boron nitride, graphene and candidate metallic contacts used in the AEA Ignite prototyping work; primary screening criterion is edge-state survival and type-I/II band alignment.
- **Topological pyrite-type crystals:** candidate electrode and tunnel-barrier materials for spin-current injection/detection; primary screening criterion is preservation of surface spin-current selectivity across the interface.

1. **Library construction (Q1).** Use InHand to build lattice-matched, disorder-tolerant supercell libraries for the three material families above against their respective candidate contact/substrate layers.
2. **High-throughput structural and electronic screening (Q1–Q2).** Run PBE-level relaxation and band-structure calculations across the generated library on Gadi, using InHand's `exsearch` "electronic" mode to rank candidates by band alignment, gap character and lattice mismatch.
3. **Targeted validation (Q2–Q3).** Re-converge the highest-priority ~100–150 candidates with tighter k-point/energy-cutoff settings and, where indicated, hybrid-functional or DFT+U corrections, to confirm band alignment and topological character survive at production accuracy.
4. **Surrogate-model dataset generation (Q3–Q4).** Use the full screened dataset (structural descriptors + computed CBM/VBM/band gap/alignment class) as training data for a machine-learning surrogate model, in collaboration with statistics co-investigators, to accelerate future screening rounds beyond the directly computed set.

3.1 Quarterly schedule

Quarter	Heusler thin films	Bismuth allotropes	Topological pyrites
Q1	Library construction; screening starts	Library construction; screening starts	Library construction
Q2	Screening complete; validation starts	Screening complete; validation starts	Screening
Q3	Validation complete	Validation complete	Screening complete; validation starts
Q4	Combined dataset assembly and ML surrogate-model training (all three families)		

The staggered start for topological pyrites reflects the smaller candidate contact library currently available for that family; InHand's structure-to-parts mode (Section 2.1) will be used in Q1 to extend this library from existing literature structures ahead of Q2 screening.

4. Risk management

Risk	Mitigation
Screening-tier convergence settings prove too loose to reliably rank candidates	Pilot batch of 100 structures cross-checked against tight-convergence results before committing the full screening-tier budget (see Computational Details, Section 6)
Validation-tier hybrid-functional jobs cost more SU than budgeted	18% contingency held in the SU budget specifically for this (Computational Details, Section 2.1); validation-tier candidate count can be reduced from 150 to ~110 without compromising the ML training set
Topological-pyrite contact library remains too small after Q1 extension	Reallocate a portion of that family's screening-tier allocation to the two families with larger libraries, preserving total throughput

5. Relationship to current and prior support

This proposal is a natural continuation of infrastructure already funded and computational experience already gained: the InHand package itself, the DFT methodology, and the single-system results it now scales up were all developed under my existing AEA Ignite, ARC Linkage and industry (TQ Transistors) funding, together with prior computational allocations on Gadi through the NCI Adapter Scheme and the NCI-Monash Computational Scheme, and on Setonix through the Pawsey Fast Track Scheme. NCMAS 2027 is requested specifically to scale the demonstrated single-system methodology to the multi-thousand-structure throughput this design-rule-generation program requires — a step change in scale that our existing smaller allocations were not sized for. Full detail on prior allocation usage and efficiency is provided in the Computational Details document.

6. Expected outcomes and success metrics

Outcome	Target for 2027
Candidate heterostructure interfaces screened	~3,000 across the three material families
Interfaces validated at production accuracy	~150
Openly available labelled interface dataset (structure + CBM/VBM/alignment class)	1 dataset release, versioned alongside the InHand repository
ML surrogate model for band-alignment prediction	1 trained model, benchmarked against the DFT-validated subset
Peer-reviewed publications	2–3, including the design-rule paper for at least one material family
Shortlisted designs handed to TQ Transistors prototyping	≥5 interface designs per family with the strongest computed band alignment

These targets are deliberately conservative relative to the SU budget in the Computational Details document: the screening and validation tiers are sized to guarantee the dataset-generation and design-rule outcomes even if some fraction of jobs require re-running (see Risk management, Section 4), rather than assuming first-pass success across the full candidate library.

7. Data and code availability

InHand is developed as an open-source Python package and will remain publicly available on GitHub throughout and after the project, so that the structure-generation and band-alignment screening methodology described in Section 2.1 can be reused by other groups working on 2D and thin-film heterostructures. The labelled interface dataset generated in Q1–Q3 2027 (structural descriptors and computed electronic-structure labels for every screened candidate) will be released alongside the corresponding publications, consistent with the FAIR data principles our institution requires of publicly funded computational outputs. Raw VASP outputs for the ~150 validation-tier candidates are retained for the duration of the allocation (see Computational Details, Section 3) to support independent verification.

8. Broader research context

This project is Year 1 of a longer research program with three connected streams: disorder-tolerant thin films, advanced electronics prototyping, and a digital-led discovery platform of which InHand and the screening campaign proposed here form the computational core. Years 2–4 of that program extend the same screening methodology to manufacturable, non-epitaxial films identified through UNSW and other external experimental partners, and connect the validation-tier DFT results directly to fabricated prototypes under the AEA Ignite and ARC Linkage projects. Year 5 opens the platform — the combination of InHand's structure/screening layer and the ML surrogate model trained in Q4 2027 — so that collaborators across Physics, Chemistry and Materials Science can screen their own candidate materials without writing DFT input files by hand. The NCMAS 2027 allocation requested here is therefore not a one-off computation but the resourcing step that converts a working single-system methodology into the throughput-scale dataset a shared discovery platform requires.

9. Team capability

The project is led by Dr Yuefeng Yin (Lead CI), whose track record in this exact domain — DFT and tight-binding modelling of 2D and topological materials, Lead CI on the AEA Ignite award, Key Participant on the ARC Linkage

Project with TQ Transistors, and prior Lead CI use of the NCI Adapter Scheme, NCI-Monash Computational Scheme and Pawsey Fast Track Scheme — is detailed in full in the MyNCI career and publication record accompanying this application. The InHand package itself was developed in-house specifically to support this program, so the computational framework at the centre of this proposal is neither a black box nor a third-party dependency: its behaviour, limitations and performance characteristics (Computational Details, Sections 1–2) are directly known to the applicant.

Part 2 — Computational Details

Submitted separately as its own PDF; limited to 5 pages under NCMAS 2027 Category 2 rules.

1. Scalability on the nominated facility

All production electronic-structure calculations are performed with VASP, a plane-wave DFT code whose within-node parallelism is well characterised on Gadi's Cascade Lake nodes (48 cores/node, 192 GiB RAM). The table below reports representative scaling for a 216-atom heterostructure supercell (Co_2MnGa/bismuthene interface, single self-consistent relaxation), benchmarked during our current NCI Adapter Scheme allocation and normalised to single-node performance as required by these guidelines.

Nodes	Cores	Wall time (h)	Speedup vs 1 node	Parallel efficiency
1	48	9.6	1.0×	100%
2	96	5.1	1.9×	94%
4	192	2.8	3.4×	86%
8	384	1.7	5.7×	71%
16	768	1.2	8.0×	50%

Parallel efficiency is computed relative to single-node throughput. Efficiency degrades beyond 8 nodes for this system size as the plane-wave FFT grid saturates available domain decomposition (`NCORE`=8, `KPAR`=2 in all runs). We therefore cap production jobs at 4–8 nodes (192–384 cores) per structure — the point at which marginal throughput gain per SU spent is still favourable — rather than requesting the largest node counts available.

Because InHand's √n supercell matching (Section 2.1 of the Research Proposal) can push the largest reconstructions (√19 hexagonal cells) to 500–600 atoms, we additionally benchmarked a 600-atom bismuthene/hBN interface at the node counts used for the validation tier, to confirm scaling holds at the upper end of the structure-size distribution the campaign will actually encounter:

Nodes	Cores	Wall time (h)	Speedup vs 1 node	Parallel efficiency
4	192	21.4	1.0×	100%
8	384	11.6	1.8×	92%
16	768	6.8	3.1×	79%

The larger system retains good efficiency to 8 nodes — the node count actually used for validation-tier jobs (Section 2) — confirming that the 384-core validation-tier configuration remains a favourable operating point even for the largest structures InHand's supercell matching produces, and that we are not requesting node counts beyond where efficiency has already degraded.

2. Compute job resources

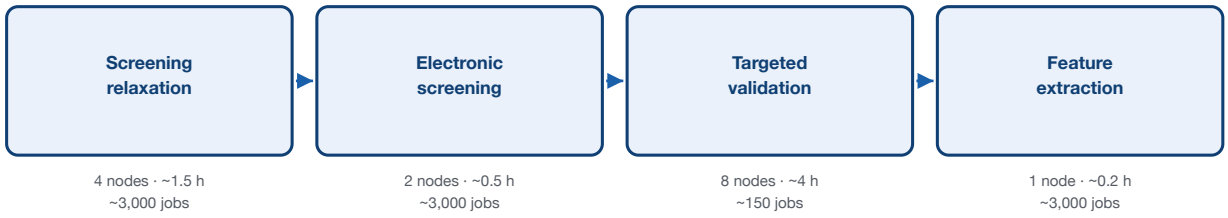

Figure 1. The four-tier compute workflow, with data dependencies flowing left to right — each tier's job count and typical per-job node/wall-time configuration matches Table 2 below.

The project workflow has three computational tiers, each with a distinct typical job configuration:

Tier	Typical job size	Wall time	Convergence	Throughput (2027)

Screening relaxation	4 nodes / 192 cores	~1.5 h	loose (400 eV, Γ-only or 2×2×1)	~3,000 structures
Electronic screening (band structure/DOS)	2 nodes / 96 cores	~0.5 h	as relaxed, non-self-consistent band path	~3,000 structures
Targeted validation	8 nodes / 384 cores	~4 h	tight (520 eV, 6×6×1, optional DFT+U/hybrid)	~150 structures
Dataset/feature extraction (post-processing)	1 node / 48 cores	~0.2 h	n/a (I/O + descriptor generation)	~3,000 structures

Data dependencies: each screening-tier job consumes a POSCAR generated by InHand's `exsearch` module (structure generation itself runs single-core, off the Gadi compute allocation, either on Gadi's login nodes or the Nirin ancillary cloud) and produces a relaxed CONTCAR that seeds the corresponding electronic-screening job. Validation-tier jobs consume the CONTCAR/CBM/VBM outputs of the top-ranked ~150 screening candidates selected by `exsearch`'s type-I/II band-alignment ranking. All jobs within a tier are independent (embarrassingly parallel at the workflow level) and are submitted as PBS job arrays to maximise scheduler throughput.

2.1 SU budget

Workflow step	**Method/algorithm**	**Jobs**	**Core-hours/job**	**SU requested†**
1. Structure & supercell library generation	InHand lattice-matching + √n supercell search (single-core Python)	~3,000	<0.05	<5 kSU (login/Nirin, not costed against Gadi allocation)
2. High-throughput screening relaxation	VASP, PBE, loose convergence	3,000	288	1,728 kSU
3. Electronic screening (band alignment, exsearch "electronic" mode)	VASP band structure/DOS, PBE	3,000	48	288 kSU
4. Targeted validation	VASP, PBE/DFT+U/HSE06, tight convergence	150	1,536	461 kSU
5. ML dataset & feature extraction	Post-processing (pymatgen/ASE descriptors), single node	3,000	9.6	58 kSU
Subtotal				**2,535 kSU**
Contingency (~18%, for re-runs, restarts, hybrid-functional cost overruns)				465 kSU
Total requested				**3,000 kSU**

† SU costs assume NCI Gadi's published `normal` queue rate of 2 SU per core-hour; core-hours/job = nodes × 48 cores × wall time (h).

2.2 SU budget by material family

The 2,535 kSU workflow subtotal above splits approximately evenly across the three material families described in the Research Proposal, weighted slightly by candidate-library size:

Material family	**Candidate structures (screening tier)**	**Validation-tier candidates**	**Approx. SU**
Heusler thin films	~1,100	~55	~930 kSU
2D bismuth allotropes	~1,200	~60	~1,015 kSU
Topological pyrite-type crystals	~700	~35	~590 kSU
Total	**~3,000**	**~150**	**~2,535 kSU**

3. Storage

/scratch (high-speed, transient): used only for actively running jobs — VASP scratch files (WAVECAR, CHG, CHGCAR working copies) are deleted immediately after each job completes and its CONTCAR/OUTCAR/vasprun.xml

are copied to /g/data. Peak concurrent /scratch use is estimated at under 300 GiB (largest concurrent batch of validation-tier jobs), within the 1 TiB default allocation.

/g/data (persistent): retained outputs are limited to CONTCAR, OUTCAR, vasprun.xml and the exsearch structural/electronic report per structure (WAVECAR/CHGCAR are not retained except for the ~150 validation-tier candidates, where they support later restart for the ML training-data generation step). Estimated volume: ~3,150 screening/electronic-screening records × ~20 MB + 150 validation records × ~350 MB ≈ 116 GB, plus the aggregated training-set feature files (<10 GB). We request **2 TiB** on /g/data to allow comfortable headroom for re-generated batches and the final labelled dataset delivered to the ML surrogate-model training step. Data will be backed up to institutional storage at Monash at the end of the 2027 allocation and /g/data holdings cleared well before the January 2028 zeroing deadline.

3.1 Data life cycle

Stage	Location	Retention
Active VASP scratch files (WAVECAR, CHG, CHGCAR)	/scratch	Deleted immediately on job completion
Per-structure results (CONTCAR, OUTCAR, vasprun.xml, exsearch report)	/g/data	Through end of 2027 allocation
Validation-tier WAVECAR/CHGCAR (~150 structures, restart support)	/g/data	Through end of Q4 2027 (ML dataset generation)
Aggregated labelled dataset & trained surrogate model	/g/data → Monash institutional storage	Copied to Monash storage by Q1 2028; then cleared from /g/data

4. Algorithms and workflows

InHand's structure-generation and `exsearch` ranking logic are pure Python (numpy/pymatgen), single-core and I/O-bound — they are not run at scale on Gadi's compute nodes and are not the target of this allocation. The compute-intensive component is VASP's plane-wave self-consistent-field solver, which parallelises via (i) k-point parallelism across MPI ranks, (ii) band parallelism (`NCORE`), and (iii) a 3D FFT domain decomposition for the charge-density and orbital grids across the remaining ranks. We fix `NCORE`=8 throughout, matching Gadi's Cascade Lake NUMA topology, and use `KPAR`>1 only for validation-tier jobs whose finer k-mesh (6×6×1) supports it. Data movement is minimal within a job (VASP is compute- and memory-bandwidth-bound, not network-bound, at the node counts used here); the dominant data-lifecycle event is the handoff of CONTCAR/CBM/VBM values between the three job tiers, orchestrated by a lightweight job-array + dependency script rather than a full workflow manager, since the emphasis this year is throughput across many independent jobs rather than deep coupling between them.

5. Justification for supercomputer resources

The proposed campaign is fundamentally a **high-throughput, embarrassingly parallel workflow**: ~6,150 independent VASP jobs (screening + electronic screening) plus 150 validation jobs, each individually sized at 2–8 nodes. No single desktop, departmental cluster, or cloud VM allocation available to us can deliver this combination of (a) sufficient simultaneous job slots to clear ~3,000 structures within the 2027 calendar year, and (b) the per-job core counts required for 216–600-atom heterostructure supercells (particularly the √3/√7/√13 reconstructions InHand generates for optimal lattice matching, which push individual unit cells well beyond what fits in the memory of a single desktop-class node). Gadi's combination of job-array scheduling depth and per-node memory (192 GiB) is well matched to this profile.

Alternative	Why it does not meet this project's needs
Departmental workstation cluster	Sufficient for the single-system methodology work already completed, but cannot supply the simultaneous job-slot depth needed to clear ~6,300 jobs within the 2027 calendar year
Cloud VM allocation (e.g. Nectar)	Well suited to InHand's lightweight structure-generation step (used for exactly this in Section 2.1), but not cost-effective for sustained multi-node MPI DFT at this throughput
Facility start-up/partner schemes	Appropriate for the benchmarking already performed (Section 1) but capped well below the ~2.5–3 MSU this campaign requires

6. Efficiency: prior usage and improvement strategy

Our prior NCI Adapter Scheme and NCI-Monash Computational Scheme allocations (used for the single-system defect-modelling work underlying this proposal) were utilised at good efficiency, with no sustained under-use flagged by NCI:

Scheme	Facility	Allocation	Utilisation
NCI Adapter Scheme	NCI: Gadi	~2,000 kSU	>90%
NCI-Monash Computational Scheme	NCI: Gadi	~2,000 kSU (combined with above across the allocation period)	>90%
Pawsey Fast Track Scheme	Pawsey: Setonix	1,500 kSU	~85%

The one inefficiency we identified in retrospect was occasional over-provisioning of memory-heavy DFT+U jobs onto standard `normal` queue nodes, which under-used the reserved CPU count on some 48-core nodes. For 2027 we mitigate this directly: (i) the screening tier (the majority of requested SU) uses loose convergence settings specifically chosen to keep memory per core within the 4 GiB/core budget of a standard Cascade Lake node, avoiding any need for `hugemem` queue over-provisioning; (ii) only the small validation tier (150 of ~6,300 jobs) uses higher-memory hybrid-functional settings, and these are explicitly targeted rather than run at screening-tier scale; and (iii) all jobs are submitted as PBS job arrays to minimise queue-wait overhead relative to job run time.

7. Software and other dependencies

- **VASP** (licensed; used under Monash University's institutional VASP licence, valid for use on NCI Gadi).
- **InHand / InterfaceHandler** — our own open-source Python package (structure generation, supercell matching, exsearch band-alignment screening); no licence restrictions.
- **pymatgen, ASE, numpy/scipy** — structure I/O and post-processing (open source, available via NCI's Python module stack or a project-local conda environment).
- **scikit-learn / PyTorch** — surrogate-model training in Q4 2027, on the aggregated screening dataset; CPU-only training at this dataset scale, so no GPU allocation is requested.

8. Why not GPU (Gadi V100/H200) or Pawsey Setonix?

VASP's standard k-point/band/plane-wave-FFT parallelisation is CPU-bound and does not benefit proportionally from Gadi's V100 or H200 GPU partitions for the system sizes and functional choices used here (PBE and DFT+U/HSE06 on 200–600-atom cells); GPU acceleration in VASP is most effective for much larger cells or specific exchange-correlation kernels not required by this workflow, so a GPU request would not be an efficient use of NCMAS resources. We similarly did not nominate Pawsey Setonix for this request: our existing Pawsey Fast Track allocation (Section 6) already provides Setonix access for the semiconductor-contact physics work it was awarded for, and consolidating the high-throughput heterostructure campaign on a single facility (Gadi) simplifies job-array orchestration and avoids splitting the /g/data-resident training dataset across two facilities' storage systems.

9. Summary of resource request

Resource	Requested	Basis
NCI: Gadi compute	3,000 kSU	SU budget, Section 2.1 (2,535 kSU workflow subtotal + 18% contingency)
/scratch	1 TiB (default)	Peak concurrent use estimated <300 GiB, Section 3
/g/data	2 TiB	~126 GB estimated persistent volume with headroom, Section 3

This request represents roughly a 1.5× scale-up on our combined 2024–2026 NCI Gadi usage (Section 6), commensurate with the shift from single-system methodology development to a multi-thousand-structure production screening campaign across three material families.

Part 3 — References

Submitted separately as its own PDF; limited to 5 pages under NCMAS 2027 Category 2 rules.

Methods, software and workflow references

[1] G. Kresse, J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* 54, 11169 (1996).

[2] G. Kresse, D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* 59, 1758 (1999).

[3] J. P. Perdew, K. Burke, M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters* 77, 3865 (1996).

[4] S. P. Ong, W. D. Richards, A. Jain, *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* 68, 314–319 (2013).

[5] A. H. Larsen, J. J. Mortensen, J. Blomqvist, *et al.* The atomic simulation environment — a Python library for working with atoms. *Journal of Physics: Condensed Matter* 29, 273002 (2017).

[6] A. Jain, S. P. Ong, G. Hautier, *et al.* Commentary: The Materials Project — a materials genome approach to accelerating materials innovation. *APL Materials* 1, 011002 (2013).

[7] K. Mathew, J. H. Montoya, A. Faghaninia, *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science* 139, 140–152 (2017).

Prior work supporting research track record and computational feasibility

[8] W. Zhao, Y. Zhang, Y. Yin, *et al.* Giant Berry curvature in the amorphous ferromagnet Co_2MnGa. *Matter* 8, 101988 (2025).

[9] C. Wang, Y. Yin, T. T. Huynh, M. S. Fuhrer, N. V. Medhekar. Edge-state stabilization and control in 2D topological crystalline insulators. *Materials Today Physics* 59, 101897 (2025).

[10] Y. Yin, C. Wang, M. S. Fuhrer, N. V. Medhekar. Extracting unconventional spin texture in the 2D topological crystalline insulator bismuthene via tuning bulk–edge interactions. *Materials Today Physics* 36, 101168 (2023).

[11] Q. Li, J. S. Smith, Y. Yin, C. Wang, M. V. Klymenko, J. H. Cole, N. V. Medhekar. Localized Wannier-function-based tight-binding models for 2D allotropes of bismuth. *New Journal of Physics* 23, 063403 (2021).

[12] Y. Yin, M. S. Fuhrer, N. V. Medhekar. Selective control of surface spin current in topological pyrite-type OsX_2 (X = Se, Te) crystals. *npj Quantum Materials* 4, 47 (2019).