MODELING COLLABORATORY FOR SUBDUCTION

TOWARD AN INTEGRATIVE COMMUNITY FRAMEWORK FOR SUBDUCTION ZONE MODELS

The Modeling Collaboratory for Subduction (MCS) is a community- and model-building effort to advance subduction zone science. The MCS group was envisioned as a core component of the Subduction Zones in Four Dimensions (SZ4D) initiative in the *Boise Subduction Zone Observatories report* (McGuire et al., 2017). Given the novelty and ambitious scope of MCS and SZ4D, their planning efforts were subsequently funded as separate RCNs. In April 2022, the MCS and SZ4D RCNs officially joined under the overarching SZ4D umbrella, and MCS will become one of the core facilities and activities of SZ4D. The MCS RCN supported a series of community workshop discussions that are summarized in detailed workshop reports (Wada et al., 2019; Dunham et al., 2020; Wada & Karlstrom, 2020; Gonnermann et al., 2021). This chapter highlights the key points that arose from these workshops, planning efforts in conjunction with the SZ4D working and integrative groups, and broad community discussions. Specifically, this chapter focuses on the structure and implementation of MCS; the detailed scientific objectives of the MCS group are woven into the individual working group chapters.

The objective of MCS is to create new kinds of physics-based models for subduction zone hazards and apply them to understand fundamental physical processes, guide instrumentation deployments, interpret observations, and assess predictability of hazards. Specifically, MCS activities will be built around the following overarching objectives:

- Construct models that link subduction zone state and long-term margin evolution to the character and predictability of event occurrence.
- Integrate observational constraints into models, while simultaneously using models to define optimal observational strategies.
- Build physics-based, predictive models for volcano, earthquake, and geomorphic systems that are spatially and temporally coupled.
- Build a diverse and equitable community of scholars who are well versed in modern modeling tools.

At a more granular level, the MCS group will work toward addressing the science questions posed by the SZ4D working groups. The individual working group chapters describe the modeling tools required to achieve their goals and identify specific areas where MCS input will be needed for success.

From the outset, MCS will integrate physics-based modeling and discovery into the observational and laboratory efforts of SZ4D, rather than following the more typical sequence in which modeling is used only after data collection. In this way, modeling will be employed not only for interpreting datasets, but in the planning and design of observational deployments. As new observations come online, data streams will be assimilated into models to assess the "state" of megathrust and volcanic systems, recognizing that long-term processes and geologic/geophysical context can play key roles in shaping individual rupture or eruptive events. MCS will employ adjoint models and physics-enabled machine learning and artificial intelligence approaches that fully leverage the new datasets being collected. In this way, we seek to transform how large-scale Earth programs are conducted and enhance what they can achieve in terms of advancing solid Earth systems science.

In the context of MCS, "model building" is a means to validate new physical descriptions, make predictions based on simplified theoretical approaches, and develop numerical models that can be used to explore the general role and interactions of fundamental processes in controlling system behavior (i.e., for the physics-based discovery of emergent phenomena). Numerical modeling can also be integrated with laboratory experiments or analog physical models to facilitate the up-scaling of laboratory results to large-scale natural systems.

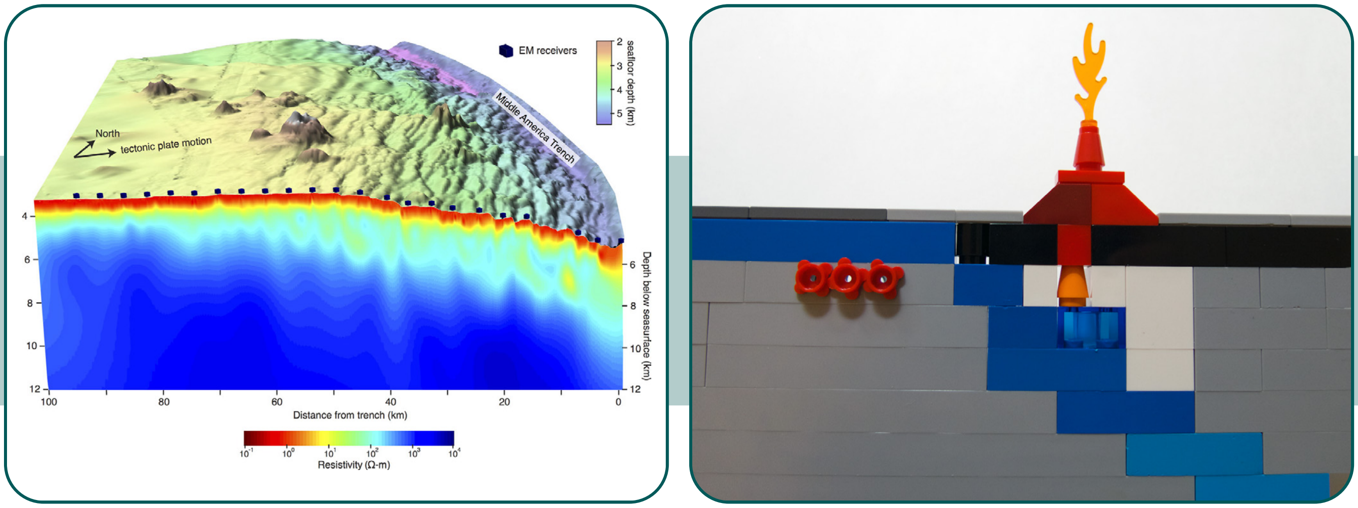
In addition, MCS seeks to build more complex, "applied," and regionally "realistic" models that can fully assimilate both structural information (e.g., from geophysical imaging and geology) and time-dependent sensor data streams (e.g., from seismometers and geodetic sensors) from global subduction zone observatories (**Figure MCS-1**). One exciting new direction in this regard is the development of adjoint models based on full or reduced-order physical models, as well as physics-enabled machine learning techniques, the combination of which have the potential to inform real-time hazard assessments alongside more traditional inversions of multi-sensor data.

Such new approaches are needed to consistently interpret constraints on the general workings of earthquakes, volcanoes, and surface processes in subduction zones; to identify knowledge gaps in our physical models; and to define optimal observational strategies to reduce uncertainties. Construction of the relevant tools will require extensive validation and verification efforts, which have the added benefit of pushing forward quantitative standards for modeling and inversion across earthquake and volcano science, generally (e.g., SCEC). On the decadal timescale, MCS will lead to the new fundamental science and operational approaches that are needed for quantifying, and possibly forecasting, earthquake, tsunami, landslide, and volcanic hazards.

The modeling envisioned by MCS will be advanced through SZ4D's ambitious plan for new observational infrastructure in Chile and complementary sites in Cascadia and Alaska. The new data streams from MegaArray, VolcArray, and SurfArray will be used to inform and test models, leading to further model refinement and possibly incorporation of additional physical processes. Moreover, it is clear that subduction zones on Earth are diverse in terms of their tectonic setting and/or current stage within their volcanic or earthquake cycles. To advance subduction zone science, it is therefore imperative to also integrate observations from other regional laboratories outside of SZ4D's primary focus sites to arrive at a globally validated physical understanding.

MCS is envisioned as a new SZ4D core facility that can serve to support the development of such a framework and provide a home for sustained interactions between modelers, experimentalists, and observationalists. Sustained efforts are crucial because we do not yet know how to assemble complete models of earthquakes, volcanoes, and landslides. MCS can provide and assimilate building blocks to test alternative physical descriptions and identify the most important processes in controlling system behavior, providing the glue between studies of subduction zones through geologic time and current activity. Moreover, MCS can support global subduction zone research communities, including the local stakeholders, as well as international and domestic observatories in different stages

Figure MCS-1. *Concept of a modular, building- block-based framework for physics-based modeling provided by the MCS to explore general physical processes and to create regionally specific subduction zones modes, for example, to interpret real-time sensor data for system "state." Left: Electric conductivity structure for Central America (from Naif et al., 2015).*



of their implementation, helping to support cross-disciplinary and cross-site collaborations. Model-based cross-validation is needed to quantitatively link insights from the Chilean observatory, as well as other international efforts, to hazard and risk settings of domestic societal concern for the United States.

Further, while any successful study of subduction zones requires an appreciation of their physical and geologic diversity, it also requires acknowledging the importance of fostering a diverse scientific community to perform these studies. MCS presents a novel opportunity to help establish computational approaches as alternative entry pathways for underserved and underrepresented communities into the geosciences. Such efforts will complement more traditional training and community-building efforts to empower computational scientists with the interdisciplinary tools they need to be successful. Together with extensive, sustainable, equitable, and coordinated outreach and teaching efforts (e.g., to enhance quantitative literacy in K–12 students and undergraduates), MCS and the computational geosciences, in general, can contribute greatly to efforts to build a better and more diverse community of geoscientists. This important theme is further explored in the BECG chapter. MCS will play a large contributing role in initiating and supporting belonging, accessibility, justice, equity, diversity, and inclusion (BAJEDI) efforts within SZ4D, as well as partner with aligned efforts such as Computational Infrastructure for Geodynamics (CIG), Community Surface Dynamics and Modeling System (CSDMS), and Cooperative Institute for Dynamic Earth Research (CIDER) domestically and internationally.

COMMUNITY-DEFINED MCS DESIGN GOALS

Through the workshops and community engagement sponsored by the MCS RCN, key design goals were defined in order for MCS to achieve the guiding objectives described above. The community emphasized the need to create new pathways for discovery, based on investments in human and computational infrastructure and that are supported over timescales longer than a typical grant cycle. The MCS RCN workshop reports describe these community-defined design goals in significant detail. Here, we highlight the common themes that crystallized from these community discussions among observationalists, experimentalists, and modelers. Specifically, the community recommends that MCS should:

- Support the development of new physics-based forward models to study geologic hazards associated with subduction zones. Model development in these efforts should be guided by the science objectives of the SZ4D working groups with priorities defined by community input.
- Support sustained exchange between computational, observational, and laboratory subduction zone scientists within SZ4D and the broader community through workshops, hackathons, and shared model building. Specifically, this scientific exchange can be realized through regional and process-focused research groups.
- Support model and modeling framework tool development and validation/verification exercises, with a mix of centralized and distributed approaches, based on continuous community input and guidance,

and supporting both community-based and PI-driven code development, as well as centralized framework efforts.

- Support the documentation and usability of codes through tutorials, cookbooks, and workflow examples to allow the use of models for both teaching and research applications.
- Support SZ4D operations by storing and disseminating data-derived products, such as community structural models, and serving as a repository for inverse and forward modeling and data analysis tools from all disciplines involved in SZ4D.
- Support access to computing resources, which will empower scientists with different backgrounds and institutional support, and broaden and democratize participation in leading-edge, data-driven, high-performance and cloud computing within the solid Earth sciences.

While striving to achieve these goals, MCS should be strategically guided by principles such as:

- Providing validated, flexible, robust, well-documented, and efficient open-source codes with inherent consideration of multi-physics, cross-scale, adjoint approaches, and uncertainty quantification.
- Embracing the guiding principles of open science and FAIR (findability, accessibility, interoperability, and reusability) data practices with standardized output formats.
- Empowering the widest and most diverse representation of the community, equitable representation of all voices and supporting active international collaboration.

For many problems a range of alternative, possibly competing, modeling approaches are needed. Thus, when possible, MCS should support modular workflows (**Figure MCS-1**) rather than a single "consensus" approach for how the physics of subduction should be modeled. Model components can then be verified and benchmarked first to ensure that codes tackle the subsystem components involved in the coupled multi-scale, multi-physics problems correctly and efficiently. Moving toward validation (i.e., ensuring that the overall physical representations - the coupled subduction models - are the right ones), codes must simulate interactions across scales and explore coupled physical processes. For this, the more general framework must be tested as widely as possible both at the SZ4D sites and through the integration of data from different regional subduction zones, laboratory experiments, and other natural laboratories.

Lastly, to be successful, MCS must ensure tight collaboration between computational and applied math experts and domain scientists, as well as close exchange between modelers, experimentalists, and observationalists. The latter includes supporting modeling and model construction by observationalists and appreciation of data analysis and laboratory experiments by modelers, in the spirit of empowering interdisciplinary research.

In summary, MCS should serve as a computational science facility for SZ4D, providing the most versatile tools possible while pursuing the goal to create physical models for subduction zone hazards. Of course, MCS will not operate in isolation. The science objectives overlap with several agencies (NSF, USGS, NASA, NOAA) and community organizations (e.g., CSDMS, CIG, Volcanology hub for Interdisciplinary

Collaboration, Tools and Resources [VICTOR]), and existing and possible future earthquake centers in the United States. There are also clear links with a number of international partners, such as the Center for Excellence in Solid Earth (ChEESE), an initiative to bring cutting-edge solid Earth high performance computing (HPC) enabled codes closer to hazard applications. MCS will strive to work collaboratively within this constellation of computational centers, leveraging existing infrastructure, while trying not to duplicate efforts.

DATA-DRIVEN COMPUTATION

The MCS is the natural home in SZ4D for data-driven computational approaches, such as deep learning and other machine learning (ML) methods that have seen significant geoscience applications in recent years. These advances are relevant to all components of SZ4D and have utility for researchers working with geophysical, geochemical, geologic, and experimental data and models. MCS combined expertise in physics-based and data-driven computation will ensure a holistic approach of SZ4D modeling and data-analysis efforts. SZ4D will both leverage existing ML techniques and develop new ML approaches in its data assimilation and modeling efforts. Specifically, MCS will utilize ML in the following ways:

1. ML data processing pipelines are adept at event and signal detection, association, location, and automatized interpretation. MCS will build and utilize these workflows to analyze real-time data collected from the on- and offshore instrumentation. Such information can be leveraged to determine seismicity distributions and fault structure as well as image magma plumbing systems and trans-crustal magma transport. MCS will also explore ML-enabled denoising, which is expected to be important for data from ocean bottom seismometers, high-rate geodesy, and other noisy data streams such as volcanic gas emissions.
2. ML-enabled workflows could be applied to rapidly process and reduce real-time multiparameter MegaArray, SurfArray, and VolcArray data for rapid, physics-based data analysis. For example, such ML workflows are now standard in hydrograph modeling. In seismology, observables such as first arriving P- and S- waves from large magnitude earthquakes can be linked to earthquake source time functions to facilitate event identification, analysis, and direct prediction of hazards such as shaking intensity and tsunami amplitudes.
3. Finally, there are opportunities to utilize and leverage the complementary strengths of ML and HPC applications. The promising results of ML methods for extracting information from large datasets can be complemented with physics-based modeling to address concerns about reproducibility and physical inconsistencies for hazard and early warning applications. For example, ML can be used to speed up forward computation and extract control parameters from complex systems. Another example is that ML could be combined with results from high-resolution simulations to reduce the required forward computations for sensitivity analysis and uncertainty quantification, and for exploring optimal experimental design.

Given that the initial planning phase of the MCS RCN was focused primarily on physics-based modeling, it will be important to carry out similar efforts for data-driven computational science applications like ML in the next phase of SZ4D. Workshops and other community-building activities are needed to identify community needs, opportunities for open-source software development, and training and educational activities.

MCS IMPLEMENTATION PLAN

To fully realize the MCS design goals described above and to support the overall SZ4D effort, we envision an MCS center that will support both centralized and distributed PI-driven modeling and science. The center will support centralized efforts spanning community code development, verification and validation exercises, training through tutorials and hackathons, and the development of community models and workflows for integration of models with data streams from SZ4D and international partners. Furthermore, the center will support distributed, PI-driven modeling and science by providing computational resources (e.g., programmer support, computational hardware allocations, planning support for focused workshops) as part of competitive proposals submitted to the SZ4D science program.

The MCS center's efforts can be divided broadly into science and operations as shown schematically in the organizational structure in **Figure MCS-2**. The center's science agenda will be set by the MCS Integrative Group in coordination with the SZ4D Collective Impact Committee; the chair of the MCS Integrative Group is also a member of the SZ4D Collective Impact Committee. These committees will set priorities for code and model development efforts, workshops, training, and other activities. The execution of these activities will be the responsibility of the MCS facility. The MCS facility staff is comprised of

- The MCS Director, a member of the SZ4D Operations Committee, who oversees the facility operations and serves as PI on proposals to support the facility.
- The MCS Program Manager, who handles day-to-day operations of the facility and handles administrative aspects of the facility such as workshop organization.
- The MCS Outreach Coordinator, who oversees capacity building activities, in coordination with the SZ4D BECG group when appropriate. It is possible that this position would be combined with the Program Manager during the initial stages of MCS until responsibilities grow beyond what a single person could manage.
- Programmers, who support centralized and PI-driven code and community model development and data integration/assimilation efforts.

MCS, in coordination with the SZ4D Collective Impact Committee and the working groups, will help to create and support the regional and process-specific focus groups described in the **Community-Driven MCS Design Goals** section above. These groups will initiate and sustain collaborations between observationalists, experimentalists, and modelers to tackle the SZ4D science questions and integrate constraints from geophysical, geological, and experimental data.

The MCS facility will also support community code development and verification and validation exercises. In many cases, these efforts will be pursued in partnership with existing and/or proposed organizations such as the CIG, CSDMS, VICTOR, and earthquake and geohazard centers

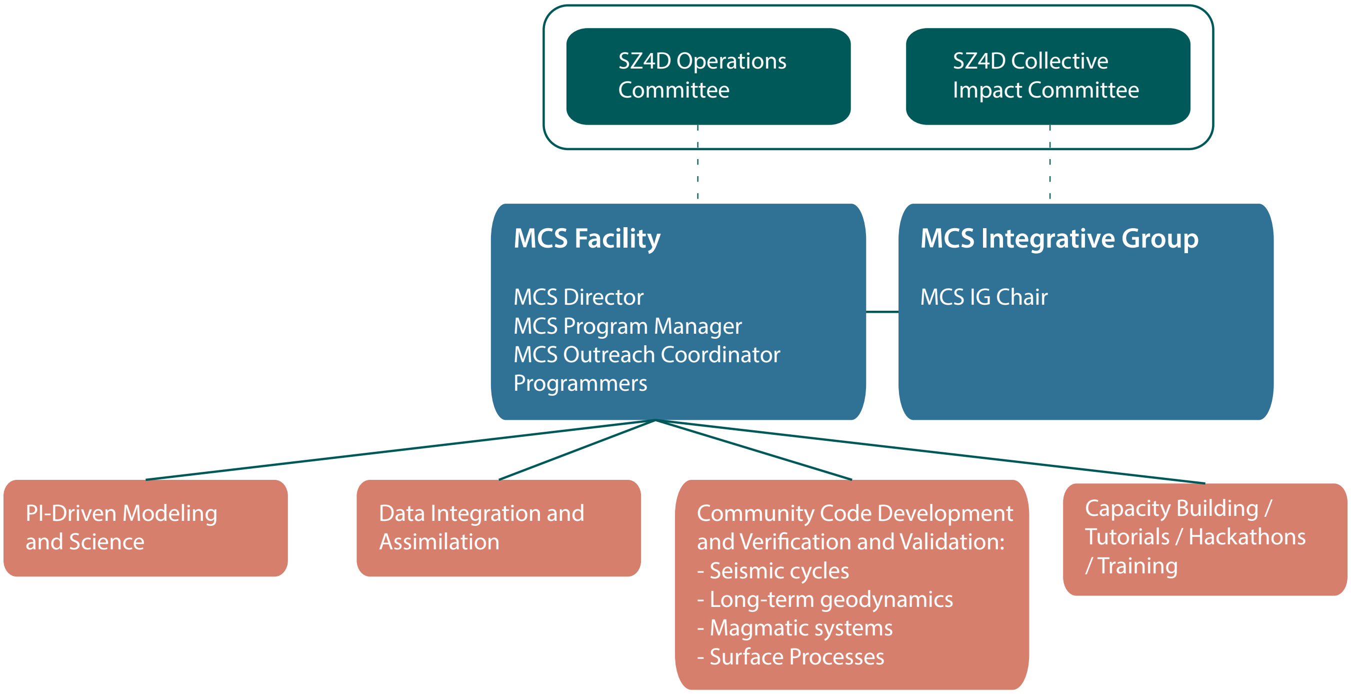


Figure MCS-2. *MCS organizational structure. Yellow boxes illustrate the two main components of the MCS: operations (MCS Facility) and science (MCS integrative group). The MCS integrative group, in coordination with the SZ4D Collective Impact Committee, will set priorities for code and model development efforts, workshops, training, and other activities performed by the MCS Facility. The orange boxes denote the activities that will be performed by the MCS Facility.*

emerging from the NSF Centers for Innovation and Community Engagement in Solid Earth Geohazards program.

As evident from the working group chapters, the current state of community modeling efforts, open-source community codes, and hence the modeling needs of each community are highly variable across disciplines. Some groups have prioritized updating widely used, but relatively simple, software packages for specific model components (e.g., forward and inverse volcano deformation models, phase equilibria calculators) to modern languages and user interfaces. Others prioritize development of open-source community codes for solving complex problems (e.g., seismic cycle modeling) or coupling existing codes into multiphysics modeling workflows (e.g., combining earthquake rupture and ground motion models with landslide and debris flow models). Across all communities there is a common need to integrate forward modeling (of either full-physics models or more computationally efficient reduced order approximations) with various data streams, through either inversions or data assimilation.

A key to this vision is for MCS to provide direct and sustained, science-driven funding for modeling and community code development activities. This vision contrasts with the operations of other existing computational facilities, such as CIG or CSDMS, which currently have minimal funding allocated to new model/code development. CIG prioritizes curating and disseminating existing software and providing training in software development best practices. This makes CIG a natural partner to MCS for certain community code development projects such as seismic cycle modeling. Furthermore, CIG maintains existing codes that can be used to address some SZ4D science questions (e.g., Aspect for long-term geodynamics and deep magma migration, PyLith for crustal deformation modeling). Similar to CIG, but focused on surface processes, the CSDMS supports computational modeling by engaging the community, encouraging the development of computational tools and protocols, and promoting education; it does not invest significant funds in new code development. In particular, the CSDMS hosts a large discoverable model repository of open-source codes compiled from the community. The CSDMS then develops and supports flexible frameworks to couple different codes together, allowing complex systems models to be assembled from modular building blocks. VICTOR is a smaller effort than either CIG or CSDMS, supporting models for ash dispersal, lava flows, and other volcano-related problems from the vent upward. There is currently no community organizational structure for developing crustal magma transport models, so the deeper focus in SZ4D'S MDE working group is thus complementary to VICTOR and potentially provides a template for advancing modeling of magmatic systems generally. Lastly, we anticipate synergies and collaborations emerging with earthquake and geohazards centers in the coming years.

In many cases, MCS-driven community code development will be pursued by the MCS facility programmers under the supervision of the MCS Director. However, it is crucial to balance this effort, which will necessarily be based upon consensus-driven science, with distributed, PI-driven model and code development. This is essential to support a diversity of ideas and approaches and to engage all interested members of the community. We envision that funding for PI-driven research will come from the NSF SZ4D Science Program, ideally through a proposal-based mechanism to obtain MCS resources such as MCS programmer time, computational hardware allocations, and planning support for focused workshops. This could be done in a similar manner to requests for seismic or geodetic support from other core facilities (e.g., the Earthscope Consortium). Resource allocation decisions would then be made by the MCS Steering Committee in coordination with the MCS Director and SZ4D CIC.

Another component of the MCS facility will be a Data Portal/Hub. Whereas the actual data from SZ4D activities will reside at the EarthScope Consortium and other facilities, the SZ4D community requires a Data Portal/Hub to house derived products such as community models, software repositories, and workflows that facilitate access to data and integration with models. We will leverage expertise from existing computational infrastructure programs and collaborate with other community centers (such as those listed above) and workflow archiving efforts to make basic data and data product infrastructures interoperable.

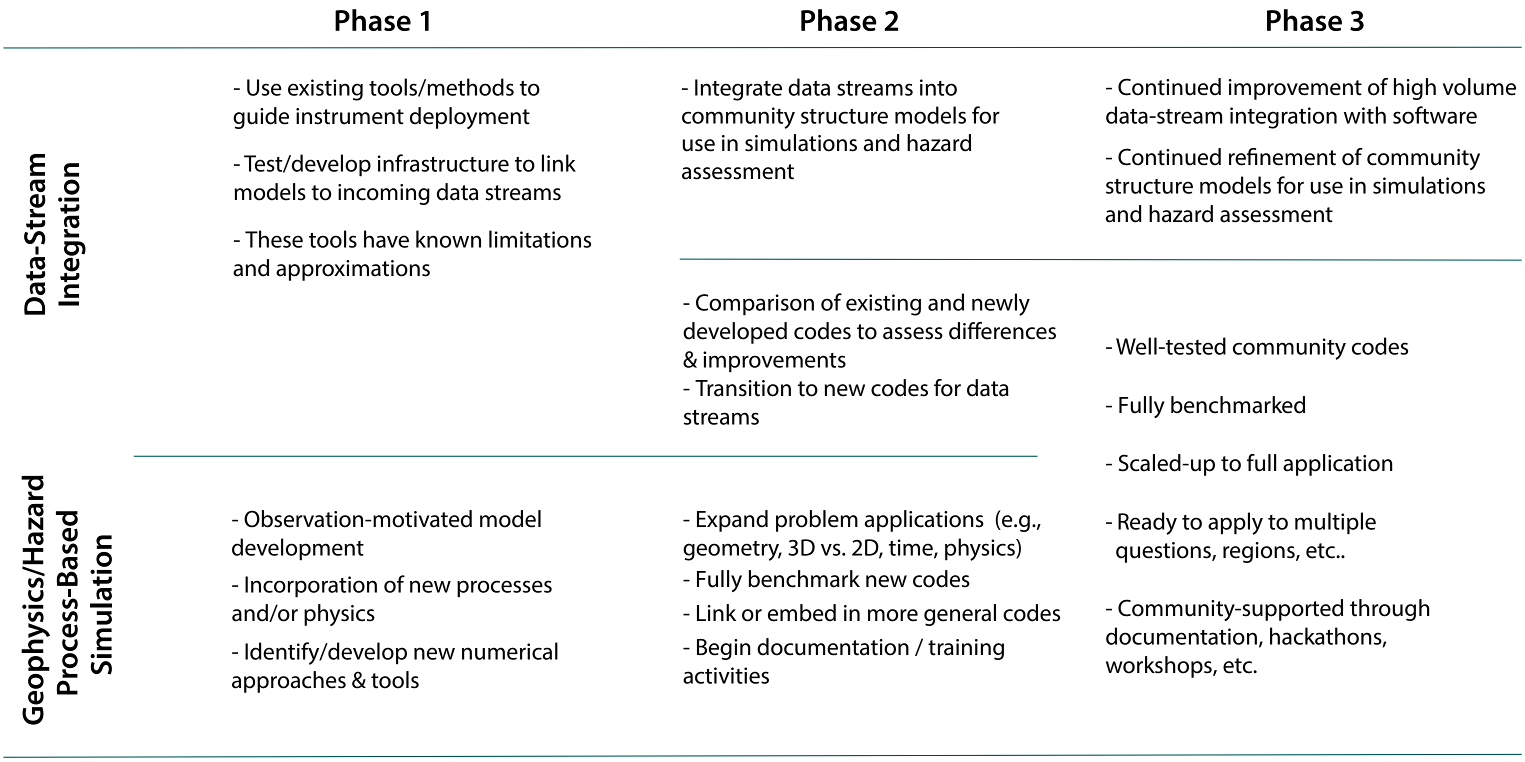
The capacity building activities, overseen by the MCS Outreach Coordinator in conjunction with the BECG group, will aim at recruiting, training, and supporting a diverse group of scientists with skills that bridge geosciences, computational science, and scientific computing. These activities will include the training programs described above but could also include fellowship programs for graduate students and postdocs. The trainees would reside with PIs and benefit from being part of a cohort coordinated through the MCS center, for example, by participating in yearly hackathons, workshops, training in software development best practices, and mentoring activities.

PHASING

Having laid out the Design Goals and Implementation Plan for MCS, we now propose a phased approach to realizing this vision. The

MCS phasing is deliberately synchronized with similar phasing in the other components of SZ4D. Phase 0, the planning phase, is culminating with the release of this report. **Figure MCS-3** summarizes the MCS phasing plan, which is framed in two parallel efforts focusing on data-stream integration and process-based modeling.

PHASE 1 focuses on the design and development of technology (including modeling and data integration software) needed to collect and interpret data from the SZ4D deployments. The MCS facility (**Figure MCS-3**) can start immediately with a Director, Program Manager (who will also serve temporarily as Outreach Coordinator), and Steering Committee. Process and regional focus groups can be initiated. Community engagement through workshops will refine the model and code development priorities beyond those identified in past MCS workshops and summarized in the MCS workshop reports and working group chapters. The MCS facility can begin hiring programmers to initiate code development, initially in a centralized manner on community codes, and broadening to support PI-driven efforts once the NSF SZ4D Science Program begins. MCS will also compile and assemble existing constraints from regional laboratories (especially Chile, Cascadia, and Alaska) and begin assembling structural and, when possible, dynamic models. These modeling efforts will be geared toward exploring optimal configurations for instrumentation arrays. Plans for the MCS Data Portal/Hub will be refined and work will begin on developing workflows for data integration and assimilation. Training efforts should also begin, which could include fellowship programs for graduate students and postdocs if funding is available.



	Phase 1	Phase 2	Phase 3
Data-Stream Integration	- Use existing tools/methods to guide instrument deployment - Test/develop infrastructure to link models to incoming data streams - These tools have known limitations and approximations	- Integrate data streams into community structure models for use in simulations and hazard assessment	- Continued improvement of high volume data-stream integration with software - Continued refinement of community structure models for use in simulations and hazard assessment
		- Comparison of existing and newly developed codes to assess differences & improvements - Transition to new codes for data streams	- Well-tested community codes - Fully benchmarked - Scaled-up to full application - Ready to apply to multiple questions, regions, etc.. - Community-supported through documentation, hackathons, workshops, etc.
Geophysics/Hazard Process-Based Simulation	- Observation-motivated model development - Incorporation of new processes and/or physics - Identify/develop new numerical approaches & tools	- Expand problem applications (e.g., geometry, 3D vs. 2D, time, physics) - Fully benchmark new codes - Link or embed in more general codes - Begin documentation / training activities	

Figure MCS-3. *Phasing diagram for MCS activities.*

PHASE 2 spans the decadal-scale observational effort. Model and code development efforts will continue, with some codes now being ready for region-specific modeling and data integration. Insights from PI-driven efforts will be incorporated into community codes. Training efforts will focus on broadening user access to include observational and laboratory scientists through tutorials, hackathons, and cookbooks. This will likely require hiring a separate MCS Outreach Coordinator. Fellowship programs should begin, if they have not already. The MCS facility will deploy the Data Portal/Hub, which will include workflow and ideally also cloud-based portals to run forward and inverse models, for both idealized and region-specific problems, and to integrate data from the SZ4D arrays. Community input based on data-integrated modeling will be used to refine hypotheses, identify knowledge gaps, and improve observational strategies.

PHASE 3 synthesizes observational data with newly developed models and experimental constraints to answer the SZ4D science questions. Community model and code development efforts will reach maturity with regard to data integration capabilities and accessibility through the Data Portal/ Hub for use by the community. PI-driven efforts will continue to explore alternative hypotheses and interpretations of SZ4D data.

We are at the cusp of achieving a new level of insight into subduction zone science and hazards, training the emerging next generation of computational subduction scientists, and elevating computational geoscience approaches to a true partnership with observational and laboratory approaches in general. Achieving this paradigm change requires bold investment and could be the beginning of a new area in solid Earth geoscience.

REFERENCES

Gonnermann, H., Anderson, K., Sisson, T., Bergantz, G., Pritchard, M., Jackson, M., Ruprecht, P., et al. (2021). *Modeling Volcano-Magmatic Systems: Crustal Magma Transport, Storage and Eruption (Draft)*, Modeling Collaboratory for Subduction, Austin TX. https://bit.ly/3nYhI9W

Dunham, E. M.,Thomas, A., Becker, T. W., Cattania, C., Hawthorne, J., Hubbard, J., Lotto, G. C., et al. (2020). *Modeling Collaboratory for Subduction RCN Megathrust Modeling Workshop Report*. EarthArXiv. https://doi.org/10.31223/X5730M

McGuire, J.J., T. Plank, et al. 2017. *The SZ4D Initiative: Understanding the Processes that Underlie Subduction Zone Hazards in 4D*. Vision Document Submitted to the National Science Foundation. The IRIS Consortium, 63 pp.

Naif, S., Key, K., Constable, S., & Evans, R. L. (2015). Water-rich bending faults at the Middle America Trench, *Geochemistry, Geophysics, Geosystems*, 16, 2582–2597. https://doi.org/10.1002/2015GC005927

Wada, I., Karlstrom, L., Arcay, D., Caricchi, L., Fulton, P., Gerya, T., Iacovino, K., et al. (2019). *Modeling Collaboratory for Subduction RCN Fluid Migration Workshop Report*. Modeling Collaboratory for Subduction, Austin, TX. https://bit.ly/3p5ia6R

Wada, I., & Karlstrom, L. (2020). Modeling fluid migration in subduction zones. *Eos*, 101. https://doi.org/10.1029/2020EO145551