Jason Gross

Email: jgross@mit.edu
Website: people.csail.mit.edu/jgross
GitHub: JasonGross
Google Scholar: QouPlrMAAAAJ

About Me

I'm a programming languages research scientist transitioning into ML and alignment. I have a working knowledge of around two dozen programming languages, and expertise in a handful (Coq (≈1M+ loc), Python (≈80k loc), Agda (≈50k loc), others). I spent my PhD on low-level cryptographic code generation, proof automation, performance engineering, and infrastructure around debugging and CI. Now I'm developing a compression-based theoretical foundation for rigorous mech interp. I'm excited about what scalable performant automation can make possible.

Experience

Special Project of ARC Theory August 2023–Present
Project Lead Berkeley, CA

- Building the first machine-checked functional-correctness proofs of mechanistic interpretability arguments about transformers in Coq (`github.com/JasonGross/neural-net-coq-interp`)
- Raised $150k & leading an interdisciplinary team of eight ML researchers and mathematicians

Machine Intelligence Research Institute (MIRI) February 2021–September 2023
Research Staff Berkeley, CA (remote)

- Performing self-directed research into fundamental programming language theory and math

Coq Development Team, INRIA June 2021–Present
Member of Core Team Nantes, France (remote)

- Reported the plurality of all-time bugs in Coq (since 2012)
- Designed and engineered a bug report minimizer for automatically producing minimized standalone test-cases and minimizing regressions in external projects tested by Coq's CI
- Researching performance issues that impact scalability of automated verification

MIT CSAIL September 2013–February 2021
PhD Researcher Cambridge, MA

- Main Project: Fiat Cryptography (`github.com/mit-plv/fiat-crypto`)
- Fiat Cryptography is a developer tool to generate proven-correct cryptographic code, with wide industry adoption, powering the plurality of secure connections in Chrome and Firefox
- Lead development of one of the world's first algorithm-level-optimizing compilers
- Collaboratively implemented the tool; wrote backends to C, Go, Java, and JSON; managed development of backends to Rust and Zig

Education

Massachusetts Institute of Technology 2013–2021
PhD in Computer Science Cambridge, MA
Advisor: Adam Chlipala
Thesis: Performance Engineering of Proof-Based Software Systems at Scale
SM Thesis: An Extensible Framework for Synthesizing Efficient, Verified Parsers

Massachusetts Institute of Technology 2009–2013
BS in Mathematics and Physics Cambridge, MA
GPA: 4.6/5

Internships

- MIRI, summer 2019: Formalized type theories, and proved properties of programs that reason about themselves
- Google, summer 2018: Worked on integration of Fiat Cryptography with BoringSSL in Chrome
- Google, summer 2016: Extended Fiat Cryptography with ECC primiatives for integration with Open Titan
- Microsoft Research, summer 2014: Collaboratively created a language for specifying input/output behavior of x86 assembly programs, verfied the input/output behavior of a number of simple programs, and improved performance of the x86proved project
- MIT CSAIL — PLV, 2012–2014: Entered a significant amount of category theory into the automated proof assistant Coq, and worked on building an interface for databases and database migration on top of category theory
- MIT CSAIL — CoCoSci, 2009–2011: Designed and managed the data collection webpage for research in categorical learning and transfer learning
- Commack High School, 2006–2009: Researched circuits over sets of natural numbers, winning 4th (2009) and 3rd (2008) place awards in mathematics at ISEF

Professional Activities

- Co-maintainer of the Fiat Cryptography project
- Co-maintainer of the homotopy type theory Coq repository (`HoTT/HoTT` on GitHub)
- Program Committee Member of ITP 2023 and CoqPL 2022
- Supervising research in formalizing correspondence of affine logic to two-player games
- Supervising research in anti-inductive utility functions
- Supervising research in performative power of predidiction markets
- Circling Facilitator at The Relateful Company
- Member of SIPB (Student Information and Processing Board)

Select Past Activities

- Particpant in MIRI Decision Theory Workshops
- Volunteer at CFAR workshops
- Instructor at MIT ESP Programs
- Instructor at Monsoon Math Camp
- President of MIT Tech Squares
- Contributor to the SIPB BarnOwn project
- Project leader for MITeX, an online interface for composing LaTeX
- TA for 6.172: Performance Engineering
- TA for 8.012: Physics I and 8.022: Physics II at the Experimental Study Group
- Participant at Cananda/USA Mathcamp

Programming Languages

- Proficient: Coq, Agda, Python, Mathematica, BASIC, TeX macro language, git, JavaScript
- Working knowledge: C, C++, OCaml, Haskell, Scheme, HTML, CSS, Perl, Bash, Java
- Basic knowledge: MATLAB, Lean, Idris, Ruby, Go, Ur/Web, x86 Assembly, SQL, Batch

Selected Presentations and Publications

[Gro+24a] Jason Gross, Rajashree Agrawal, Thomas Kwa, Euan Ong, Chun Hei Yip, Alex Gibson, Soufiane Noubir, and Lawrence Chan. *Compact Proofs of Model Performance via Mechanistic Interpretability*. accepted to The Thirty-Eighth Annual Conference on Neural Information Processing Systems. Dec. 2024. DOI: 10.48550/arxiv.2406.11779. arXiv: 2406.11779.

[Yip+24] Chun Hei Yip, Rajashree Agrawal, Lawrence Chan, and Jason Gross. *Modular addition without black-boxes: Compressing explanations of MLPs that compute numerical integration*. Dec. 2024. arXiv: 2412.03773 [cs.LG]. URL: https://arxiv.org/abs/2412.03773.

[Wu+24] Wilson Wu, Louis Jaburi, Jacob Drori, and Jason Gross. *Unifying and Verifying Mechanistic Interpretations: A Case Study with Group Operations*. Oct. 2024. DOI: 10.48550/arxiv.2410.07476. arXiv: 2410.07476 [cs.LG]. URL: https://arxiv.org/abs/2410.07476.

[Gro24a] Jason Gross. *Short Formal Proofs of Transformers via Mechanistic Interpretability*. Presented at ILIAD Conference, Berkeley, California. Aug. 2024.

[Gro+24b] Jason Gross, Andres Erbsen, Jade Philipoom, Rajashree Agrawal, and Adam Chlipala. "Towards a Scalable Proof Engine: A Performant Prototype Rewriting Primitive for Coq". In: *Journal of Automated Reasoning* 68.3 (Aug. 2024), p. 19. ISSN: 1573-0670. DOI: 10.1007/s10817-024-09705-6. arXiv: 2305.02521 [cs.PL].

[YAG24] Chun Hei Yip, Rajashree Agrawal, and Jason Gross. *ReLU MLPs Can Compute Numerical Integration: Mechanistic Interpretation of a Non-linear Activation*. accepted to ICML 2024 Workshop on Mechanistic Interpretability. June 2024. URL: https://openreview.net/forum?id=rngMb1wDOZ.

[Gro24b] Jason Gross. *Guarantees-Driven Mechanistic Interpretability: Formal Proof Size as a Metric for Mechanistic Detail of Understanding*. Presented at FAR AI's weekly seminar. Feb. 2024.

[Gro23] Jason Gross. *MetaCoq Quotation: Partial Work Towards Löb's Theorem*. Presented remotely to the Gallinette team in Nantes at an informal workshop on meta-programming and modal type theories with native quotation operations. Oct. 2023.

[Kue+23] Joel Kuepper, Andres Erbsen, Jason Gross, Owen Conoly, Chuyue Sun, Samuel Tian, David Wu, Adam Chlipala, Chitchanok Chuengsatiansup, Daniel Genkin, Markus Wagner, and Yuval Yarom. "CryptOpt: Verified Compilation with Random Program Search for Cryptographic Primitives". In: *PLDI'23: Proceedings of the 44th ACM SIGPLAN Conference on Programming Language Design and Implementation*. Distinguished Paper Award. Orlando, FL, USA, June 2023. arXiv: 2305.19586. URL: http://adam.chlipala.net/papers/CryptoptPLDI23/.

[GE22] Jason Gross and Andres Erbsen. *10 Years of Superlinear Slowness in Coq*. Presented at The Coq Workshop 2022. Aug. 2022. URL: https://jasongross.github.io/papers/2022-superlinear-slowness-coq-workshop.pdf.

[Gro+22a] Jason Gross, Andres Erbsen, Jade Philipoom, Miraya Poddar-Agrawal, and Adam Chlipala. "Accelerating Verified-Compiler Development with a Verified Rewriting Engine". In: *Proceedings of the 13th International Conference on Interactive Theorem Proving (ITP 2022)*. Ed. by June Andronick and Leonardo de Moura. Vol. 237. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Aug. 2022, 17:1–17:18. ISBN: 978-3-95977-252-5. DOI: 10.4230/LIPIcs.ITP.2022.17. eprint: 2205.00862. URL: https://jasongross.github.io/papers/2022-rewriting-itp.pdf.

[Gro+22b] Jason Gross, Théo Zimmermann, Miraya Poddar-Agrawal, and Adam Chlipala. "Automatic Test-Case Reduction in Proof Assistants: A Case Study in Coq". In: *Proceedings of the 13th International Conference on Interactive Theorem Proving (ITP 2022)*. Ed. by June Andronick and Leonardo de Moura. Vol. 237. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Aug. 2022, 18:1–18:18. ISBN: 978-3-95977-252-5. DOI: 10.4230/LIPIcs.ITP.2022.18. URL: https://jasongross.github.io/papers/2022-coq-bug-minimizer-itp.pdf.

[Gro21a] Jason S. Gross. "Performance Engineering of Proof-Based Software Systems at Scale". PhD Thesis. Massachusetts Institute of Technology, Feb. 2021. URL: https://jasongross.github.io/papers/2021-JGross-PhD-EECS-Feb2021.pdf.

[Gro21b] Jason Gross. *A Limited Case for Reification by Type Inference*. Presented at The Seventh International Workshop on Coq for Programming Languages (CoqPL'21). Jan. 2021. URL: https://jasongross.github.io/papers/2021-reification-by-type-inference-coqpl.pdf.

[Pit+20] Clément Pit–Claudel, Peng Wang, Benjamin Delaware, Jason Gross, and Adam Chlipala. "Extensible Extraction of Efficient Imperative Programs with Foreign Functions, Manually Managed Memory, and Proofs". In: *Proceedings of the 9th International Joint Conference on Automated Reasoning (IJCAR'20)*. Ed. by Nicolas Peltier and Viorica Sofronie-Stokkermans. Paris, France: Springer International Publishing, June 2020, pp. 119–137. ISBN: 978-3-030-51054-1. DOI: 10.1007/978-3-030-51054-1_7.

[Erb+19] Andres Erbsen, Jade Philipoom, Jason Gross, Robert Sloan, and Adam Chlipala. "Simple High-Level Code For Cryptographic Arithmetic – With Proofs, Without Compromises". In: *Proceedings of the 40th IEEE Symposium on Security and Privacy (S&P'19)*. May 2019. DOI: 10.1145/3421473.3421477. URL: https://jasongross.github.io/papers/2019-fiat-crypto-ieee-sp.pdf.

[Gro18] Jason Gross. *Presentation Proposal for Teaching Your Rooster to Crow in C*. Presented at The Coq Workshop 2018. July 2018. URL: https://jasongross.github.io/presentations/coq-workshop-2018/coq-workshop-proposal-notations.pdf.

[GEC18] Jason Gross, Andres Erbsen, and Adam Chlipala. "Reification by Parametricity: Fast Setup for Proof by Reflection, in Two Lines of Ltac". In: *Proceedings of the 9th International Conference on Interactive Theorem Proving (ITP'18)*. Ed. by Jeremy Avigad and Assia Mahboubi. Cham: Springer International Publishing, July 2018, pp. 289–305. ISBN: 978-3-319-94821-8. DOI: 10.1007/978-3-319-94821-8_17. URL: https://jasongross.github.io/papers/2018-reification-by-parametricity-itp-camera-ready.pdf.

[Chl+17] Adam Chlipala, Benjamin Delaware, Samuel Duchovni, Jason Gross, Clément Pit–Claudel, Sorawit Suriyakarn, Peng Wang, and Katherine Ye. "The End of History? Using a Proof Assistant to Replace Language Design with Library Design". In: *Proceedings of the The 2nd Summit oN Advances in Programming Languages (SNAPL'17)*. Ed. by Benjamin S. Lerner, Rastislav Bodík, and Shriram Krishnamurthi. Vol. 71. Leibniz International Proceedings in Informatics (LIPIcs). Asilomar, CA, USA: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, May 2017, 3:1–3:15. ISBN: 978-3-95977-032-3. DOI: 10.4230/LIPIcs.SNAPL.2017.3. URL: https://jasongross.github.io/papers/FiatSNAPL17.pdf.

[Bau+17] Andrej Bauer, Jason Gross, Peter LeFanu Lumsdaine, Michael Shulman, Matthieu Sozeau, and Bas Spitters. "The HoTT Library: A Formalization of Homotopy Type Theory in Coq". In: *Proceedings of the 6th ACM SIGPLAN Conference on Certified Programs and Proofs*. CPP 2017. Paris, France: ACM, Jan. 2017, pp. 164–172.

ISBN: 978-1-4503-4705-1. DOI: 10.1145/3018610.3018615. eprint: 1610.04591. URL: https://jasongross.github.io/papers/2017-HoTT-formalization.pdf.

[Gro16] Jason Gross. *The HoTT/HoTT Library in Coq: Designing for Speed*. Presented at The 5th International Congress on Mathematical Software (ICMS 2016). July 2016. URL: https://jasongross.github.io/presentations/icms-2016/hott-hott-and-category-coq-experience.pdf.

[Gro15a] Jason Gross. "An Extensible Framework for Synthesizing Efficient, Verified Parsers". MA thesis. Massachusetts Institute of Technology, Sept. 2015. URL: https://jasongross.github.io/papers/2015-jgross-thesis.pdf.

[Del+15] Ben Delaware, Clément Pit–Claudel, Jason Gross, and Adam Chlipala. "Fiat: Deductive Synthesis of Abstract Data Types in a Proof Assistant". In: *Proceedings of the 42nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'15)*. Jan. 2015. DOI: 10.1145/2775051.2677006. URL: https://jasongross.github.io/papers/2015-adt-synthesis.pdf.

[Gro15b] Jason Gross. *Coq Bug Minimizer*. Presented at The First International Workshop on Coq for PL (CoqPL'15). Jan. 2015. URL: https://jasongross.github.io/papers/2015-coq-bug-minimizer.pdf.

[TG15] Tobias Tebbi and Jason Gross. *A Profiler for Ltac*. Presented at The First International Workshop on Coq for PL (CoqPL'15). Jan. 2015. URL: https://jasongross.github.io/papers/2015-ltac-profiler.pdf.

[Gro14a] Jason Gross. *Presentation: Input, Output, and Automation in x86 Proved*. Presented at Microsoft Research, Cambridge, UK. Aug. 2014. URL: https://jasongross.github.io/presentations/msr-2014-final-talk/input-output-and-automation-in-x86proved.pdf.

[GCS14] Jason Gross, Adam Chlipala, and David I. Spivak. "Experience Implementing a Performant Category-Theory Library in Coq". In: *Proceedings of the 5th International Conference on Interactive Theorem Proving (ITP'14)*. Ed. by Gerwin Klein and Ruben Gamboa. Cham: Springer International Publishing, July 2014, pp. 275–291. ISBN: 978-3-319-08970-6. DOI: 10.1007/978-3-319-08970-6_18. eprint: 1401.7694. URL: https://jasongross.github.io/papers/category-coq-experience-itp-submission-final.pdf.

[Gro14b] Jason Gross. *Presentation Proposal for Three Neat Tricks in Coq 8.5*. Presented at the 6th Coq Workshop. Apr. 2014. URL: https://jasongross.github.io/presentations/coq-workshop-2014/coq-workshop-proposal-tactics-in-terms.pdf.

[Gro14c] Jason Gross. *Jason Gross' Wishlist for Coq*. Jan. 2014. URL: https://jasongross.github.io/presentations/coq-8.6-wishlist/jgross-coq-8-6-wishlist-no-pause.pdf.

[Gro14d] Jason Gross. *POPL: Minute Madness: Category Theory in Coq, and Program Synthesis*. Presented at the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'14). Jan. 2014. URL: https://jasongross.github.io/presentations/popl-2014-minute-madness/jason-gross-minute-madness.pdf.

[Gro13a] Jason Gross. *CSAIL Student Workshop 2013: Computational Higher Inductive Types: Computing with Custom Equalities*. Presented at the 2014 MIT CSAIL Student Workshop. Oct. 2013. URL: https://jasongross.github.io/presentations/csw-2013/jgross-presentation-no-pause.pdf.

[Gro13b] Jason Gross. *Building Database Management on top of Category Theory in Coq*. Presented as a student talk at the 40th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'13). Jan. 2013. URL: https://jasongross.github.io/presentations/popl-2013/jgross-student-talk.pdf.

[Gro13c] Jason Gross. *POPL: Minute Madness: Database Management on top of Category Theory in Coq: Category of Relational Schemas = Category of Categories*. Presented at the 40th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'13). Jan. 2013. URL: https://jasongross.github.io/presentations/popl-2013/minute-madness.pdf.

[Lak+11] Brenden M. Lake, Ruslan Salakhutdinov, Jason Gross, and Joshua B. Tenenbaum. "One shot learning of simple visual concepts". In: *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. 2011. URL: https://jasongross.github.io/papers/LakeEtAl2011CogSci.pdf.