Massachusetts Institute of Technology
Office: 32, Vassar Street, 32-G778
Cambridge, MA 02139, USA
+1 (917) 780 9232
charithm@mit.edu
http://charithmendis.com/

Charith Mendis

Research Interests and Vision

My broad research interests are in **compilers, program analysis and machine learning.** I envision a future around **end-to-end learned compiler optimizations** that can generate state-of-the-art performant code, while making the compiler future-proof and easier to maintain. To this end, I have developed **Ithemal**, the first learned compiler cost model and **Vemal**, the first end-to-end learned auto-vectorizer, both of which outperform state-of-the-art hand-crafted heuristics and algorithms with minimal human intervention.

Education

2015–2020 expected **Ph.D. in Computer Science**, *Massachusetts Institute of Technology*, USA.
- ○ Thesis: Towards Automated Construction of Compiler Optimizations
- ○ Advisor: Saman Amarasinghe

2013–2015 **S.M. in Computer Science**, *Massachusetts Institute of Technology*, USA.
- ○ Thesis: Helium: Lifting High-Performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code
- ○ Advisor: Saman Amarasinghe
- ○ Awards: **William A. Martin Memorial Thesis Award for outstanding SM thesis in CS**

2008–2013 **B.Sc.(Eng) in Electronics and Telecommunication Engineering**, *University of Moratuwa*, Sri Lanka.
- ○ Thesis: Application of Machine Learning Techniques for Currency Note Validation
- ○ Awards: Ranked 1st out of 650 students (Gold Medalist). Best GPA ever to-date (4.17/4.20).

1999-2007 **G.C.E.(O/L) and (A/L)**, *Royal College Colombo 7*, Sri Lanka.
- ○ Ranked 1st in Sri Lanka out of 40,000 students at G.C.E.(A/L) examination held in 2007.
- ○ Ranked 2nd in Sri Lanka out of 380,000 students at G.C.E.(O/L) examination held in 2004.

Experience

Massachusetts Institute of Technology, Cambridge

Research Assistant **Learned Compiler Optimizations (2018-present)**
Developed the first learned compiler cost model, *Ithemal* [Mendis et al., ICML 2019], that predicts throughput of basic blocks using deep neural networks. Ithemal reduces the error of prediction by half compared to hand-written tools. Developed *Vemal* [Mendis et al., NeurIPS 2019], the first end-to-end learned policy for compiler auto-vectorization which outperforms hand-crafted heuristics. As a result of a recent collaboration with the *Google Compiler Research Team*, Ithemal has already been reimplemented and used inside *Google* as part of their CPU performance modeling effort.

Solver-aided Auto-Vectorization (2017-2018)

Developed a novel superword level parallelism based auto-vectorization framework, goSLP [Mendis & Amarasinghe, OOPSLA 2018], with certain optimality guarantees. goSLP uses integer linear programming to perform pairwise statement packing optimally and dynamic programming to select the best lane assignment of vectors.

Automatic Program Rejuvenation (2013-present)

Developed *Helium* [Mendis et al., PLDI 2015], a system which dynamically analyzes the behavior of stripped x86 binaries and lifts stencil kernels to a high level language implementation in *Halide*. The users can then recompile the lifted kernels to produce fresh rejuvenated and better performing binaries. Developed a compiler optimization pass, *Revec* [Mendis et al., CC 2019], that statically analyzes hand-vectorized codes and rejuvenates its performance by generating code targeting newer and wider vector instruction sets.

Optimizing Graph Algorithms (2016-2017)

Developed a novel vertex reordering scheme and graph partitioning scheme that enables state-of-the-art performance in commonly used graph algorithms [Zhang et al., BigData 2017].

Microsoft Research, Redmond

Research Intern | **Fast Speech Decoding (2015-2016)**
Developed a novel synchronization-free parallel algorithm and graph partitioning based load balancing scheme to improve speech decoding performance [Mendis et al., ICASSP 2016].

Awards and Honors

2019	Best Paper Award	*ML for Systems Workshop @ISCA 2019*
2017	Best Student Paper Award	*IEEE International Conference on Big Data (BigData)*
2015	William A. Martin Memorial Thesis Award for outstanding SM thesis in CS	*MIT*
2013	MIT Energy Initiative Fellowship	*MIT*
2012	Sri Lanka Telecom Gold Medal	*University of Moratuwa, Sri Lanka*
2008 - 12	Awards for highest GPA in each individual year	*University of Moratuwa, Sri Lanka*
2009, 2010	IBM Engineering Scholarship for Best Academic Performance	*University of Moratuwa, Sri Lanka*
2009, 2010	Sri Lanka Telecom Scholarship for Best Academic Performance	*University of Moratuwa, Sri Lanka*
2009	Joint 8th in the world for Management Accounting - Decision Management	*CIMA*
2008	Joint 3rd in the world for Management Accounting - Performance Evaluation	*CIMA*
2008	Mahapola Merit Higher Education Scholarship	*Government of Sri Lanka*
2008	Silver Medal at Sri Lanka Physics Olympiad	*Institute of Physics, Sri Lanka*
2008	Student Peace Ambassador representing Sri Lanka	*Singapore High Commission in Sri Lanka*
2008	People's Bank and Dialog GSM scholarships for G.C.E.(A/L) performance	
2007	Special written commendation by the president of Sri Lanka for exemplary performance at both G.C.E.(A/L) and G.C.E.(O/L) examinations	
2007	1st in Sri Lanka at G.C.E.(A/L) examination (out of 40,000 students)	*Sri Lanka*
2007	Turnour Prize - Highest Academic Honor	*Royal College, Sri Lanka*
2006	2nd in Sri Lanka at Australian National Chemistry Quiz	*Sri Lanka*
2004	People's Bank and Dialog GSM scholarships for G.C.E.(O/L) performance	
2004	2nd in Sri Lanka at G.C.E.(O/L) examination (out of 380,000 students)	*Sri Lanka*

▬▬▬ Publications

Published 6 first-author papers including papers in both top programming languages venues such as PLDI and OOPSLA as well as in top machine learning venues such as ICML and NeurIPS during my PhD. Received 1 best paper award, 1 best student paper award and an outstanding SM thesis award.

Conference Publications

NeurIPS 2019.
<u>Charith Mendis</u>, Cambridge Yang, Yewen Pu, Saman Amarasinghe and Michael Carbin. Compiler Auto-Vectorization with Imitation Learning. 32nd *Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

IISWC 2019.
Yishen Chen, Ajay Brahmakshatriya, <u>Charith Mendis</u>, Alex Renda, Eric Atkinson, Ondrej Sykora, Saman Amarasinghe and Michael Carbin. BHive: A Benchmark Suite and Measurement Framework for Validating x86-64 Basic Block Performance Models. *International Symposium on Workload Characterization (IISWC)*, 2019.

ICML 2019.
<u>Charith Mendis</u>, Alex Renda, Saman Amarasinghe and Michael Carbin. Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks. 36th *International Conference on Machine Learning (ICML)*, 2019. [**Best Paper Award - ML for Systems Workshop @ISCA 2019**].

CC 2019.
<u>Charith Mendis</u>*, Ajay Jain*, Paras Jain and Saman Amarasinghe. Revec: Program Rejuvenation through Revectorization. *International Conference on Compiler Construction (CC)*, 2019.

OOPSLA 2018.
<u>Charith Mendis</u> and Saman Amarasinghe. goSLP: Globally Optimized Superword Level Parallelism Framework. *Object-Oriented Programming, Systems, Languages & Applications (OOPSLA)*, 2018.

BigData 2017.
Yunming Zhang, Vladimir Kiriansky, <u>Charith Mendis</u>, Matei Zaharia and Saman Amarasinghe. Making Caches Work for Graph Analytics. *IEEE International Conference on Big Data (BigData)* [**Best Student Paper Award**], 2017.

ICASSP 2016.
<u>Charith Mendis</u>, Jasha Droppo, Saeed Maleki, Madanlal Musuvathi, Todd Mytkowicz and Geoffrey Zweig. Parallelizing WFST Speech Decoders. *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, 2016.

PLDI 2015.
<u>Charith Mendis</u>, Jeffrey Bosboom, Kevin Wu, Shoaib Kamil, Jonathan Ragan-Kelley, Sylvain Paris, Qin Zhao and Saman Amarasinghe. Helium: lifting high-performance stencil kernels from stripped x86 binaries to Halide DSL code. *Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, 2015.

Preprints

Yishen Chen, <u>Charith Mendis</u>, John Feser, Armando Solar-Lezama and Saman Amarasinghe. Turbo Boosting Program Synthesis with Code Generation. *Under Submission*, 2019.

Theses

Charith Mendis. Helium: Lifting High-Performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code. *SM Thesis, Massachusetts Institute of Technology* [**William A. Martin Memorial Thesis Award**], 2015.

(* equal contribution).

Press

Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks

- ○ **i-programmer.** How Fast Does This Code Run? (`https://bit.ly/2Xhr2Xl`)
- ○ **MIT News.** Tool predicts how fast code will run on a chip (`https://bit.ly/3bQOOP5`)
- ○ **Register UK.** Finally, a good use for AI (`https://bit.ly/3g7As05`)

Helium: Lifting High-Performance Stencil Kernels from Stripped x86 Binaries to Halide DSL

- ○ **Fortune News.** MIT researchers have found a way to make rotten software fresh again (`http://for.tn/1HTBO6d`)
- ○ **MIT News.** Computer program fixes old code faster than expert engineers (`http://bit.ly/1KAWSOP`)
- ○ **Computer Business Review.** Quick fix for bit-rot found in new Helium programme (`http://bit.ly/1LP8TkL`)

Teaching and Mentorship

Research Supervision (Undergraduate).
- ○ Zach Holbrook (2020) - Generartive Models for Programs (MEng)
- ○ Tarindu Jayatilaka (2019) - Learning a cost model for LLVM IR
- ○ Zachary Holbrook (2019) - Generative model for DSL programs
- ○ Hoi Wai Yu (2018) - Learning a classifier to select vectorization strategies
- ○ Ajay Jain (2017) - *Revec*: program rejuvenation through revectorization [CC 2019]
- ○ Chathura Widanage, Rukshan Viduranga, Yassiru Kassapa and Lasantha Ekanayake (2016) - Extending Helium to handle non-linear filters

Research Supervision (Junior PhD students).
- ○ Yishen Chen and Ajay Brahmakshatriya (2019) - A Benchmark suite for validating x86-64 performance models [IISWC 2019]

Winter 2017 **Instructor, Workshop on Functional Programming**, *University of Moratuwa.*
Prepared new material including slides and coding assignments, taught the class for more than 100 Sophomore and Junior undergraduates.

Fall 2015 **Teaching Assistant, 6.172 Performance Engineering of Software Systems**, *Massachusetts Institute of Technology.*
Prepared and play tested projects and weekly homework assignments, prepared and moderated quizzes, held weekly recitations and office hours.

Spring 2013 **Graduate Lab Assistant**, *University of Moratuwa.*
Prepared laboratory assignments for Digital Electronics and Advanced Electronics courses, supervised weekly lab experiments.

Service

Program Committee.
- ○ European Conference on Object-Oriented Programming (ECOOP) Doctoral Symposium (2019)

Externel Review Committee.
- ○ International Conference on Parallel Architectures and Compilation Techniques (PACT) (2020)

Artifact Evaluation Committee.
- ○ Code Generation and Optimization (CGO) Conference (2020)

Graduate Admissions Committee, *Massachusetts Institute of Technology*.
Reviewed and recommended applications for admission into MIT's PhD program (2017, 2018)

Journal / Conference Reviewing.
- NeurIPS : Neural Information Processing Systems (2020)
- JPDC : Journal of Parallel and Distributed Computing (2019)
- PACT: Parallel Architectures and Compilation Techniques (2019)
- TACO: Transactions on Architecture and Code Optimization (2019)
- ASPLOS: Architectural Support for Programming Languages and Operating Systems (2019)
- SPAA: Symposium on Parallelism in Algorithms and Architectures (2018)
- CGO: Code Generation and Optimization (2016, 2017, 2018)

Talks

Modernizing Compiler Technology using Machine Learning.

ADA Center	May 2020
Microsoft Research, Redmond	Apr. 2020
University of Southern California, Special Seminar	Apr. 2020
University of Illinois at Urbana-Champaign (UIUC), Special Seminar	Apr. 2020
UC Berkeley, Special Seminar	Mar. 2020
Google Brain, CA	Feb. 2020
MIT Fast Code Seminar	Nov. 2019
Microsoft Research, Redmond	Nov. 2019

Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks.

International Conference on Machine Learning (ICML)	Jun. 2019
ML for Systems Workshop (co-located with ISCA 2019)	Jun. 2019

Revec: Program Rejuvenation through Revectorization.

International Conference on Compiler Construction (CC)	Feb. 2019

goSLP: Globally Optimized Superword Level Parallelism Framework.

LLVM Seminar, MIT	Aug. 2019
Reservoir Labs, New York	Feb. 2019
Object Oriented Programming, Systems, Languages and Applications (OOPSLA)	Nov. 2018
LLVM developer meeting. San Jose (poster)	Oct. 2018
MIT PL Seminar	Mar. 2018

Helium: Lifting High-performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code.

MIT PL Offsite	May 2016
University of Moratuwa, Sri Lanka	Jan. 2016
MIT Graphics Group Lunch	Oct. 2015
Microsoft Research, Redmond	Jun. 2015
Programming Languages Design and Implementation (PLDI)	Jun. 2015
MIT PL Seminar	Jun. 2015

References

Saman Amarasinghe
Professor and Associate Department Head
Department of Electrical Engineering and
Computer Science
Massachusetts Institute of Technology
saman@csail.mit.edu

Michael Carbin
Assistant Professor
Department of Electrical Engineering and
Computer Science
Massachusetts Institute of Technology
mcarbin@csail.mit.edu

Armando Solar-Lezama
Associate Professor
Department of Electrical Engineering and
Computer Science
Massachusetts Institute of Technology
asolar@csail.mit.edu

Madanlal Musuvathi
Principal Researcher
Partner Research Manager
Microsoft Research, Redmond
madanm@microsoft.com