

# SysML Conference

**February 15 & 16, 2018  
Stanford, CA**

SysML is a new conference targeting research at the intersection of systems and machine learning. The conference aims to elicit new connections amongst these fields, including identifying best practices and design principles for learning systems, as well as developing novel learning methods and theory tailored to practical machine

[TOP](#)

learning workflows.

The inaugural SysML conference will be held on Feb 15th and 16th, 2018, in Stanford, California. The first day will be a closed-door event among the SysML Program Committee to discuss the future of the conference. The second day will be a public event with invited and contributed talks and poster presentations.

Please note our Conference Code of Conduct.

## Schedule

Closed-Door Workshop 2/15	Main Conference 2/16
<b>SysML Conference: Friday, February 16</b>	
<b>Location: Paul Brest Hall</b>	
8:00 am - 9:00 am	Breakfast and Registration
9:00 am - 9:15 am	Opening Remarks ( <a href="https://youtu.be/eXdq58JqJKc">https://youtu.be/eXdq58JqJKc</a> ): Ameet Talwalkar
<b>Session I</b> (moderator: Virginia Smith)	
9:15 am - 9:55 am	Invited talk: SysML: Perspectives and Challenges ( <a href="https://youtu.be/4inlBmY8dQI">https://youtu.be/4inlBmY8dQI</a> ), Michael I. Jordan
9:55 am - 10:05 am	Contributed talk: TVM: End-to-End Compilation Stack for Deep Learning ( <a href="https://youtu.be/l1APhISjVjs">https://youtu.be/l1APhISjVjs</a> ), Tianqi Chen
10:05 am - 10:15 am	Contributed talk: Robust Gradient Descent via Moment Encoding with LDPC Codes ( <a href="https://youtu.be/QLPj6aQmXFw">https://youtu.be/QLPj6aQmXFw</a> ), Arya Mazumdar

TOP

10:15 am - 10:25 am	Contributed talk: Analog electronic deep networks for fast and efficient inference ( <a href="https://youtu.be/8t0Yunt5kE4">https://youtu.be/8t0Yunt5kE4</a> ), Jonathan Binas
10:25 am - 10:50 pm	Coffee Break
<b>Session II</b> (moderator: Virginia Smith)	
10:50 am - 11:30 am	Invited talk: Hardware for Deep Learning ( <a href="https://youtu.be/zDBF0xwQW-0">https://youtu.be/zDBF0xwQW-0</a> ), Bill Dally
11:30 am - 11:40 am	Contributed talk: YellowFin: Adaptive Optimization for (A)synchronous Systems ( <a href="https://youtu.be/Etp3yZgflqc">https://youtu.be/Etp3yZgflqc</a> ), Ioannis Mitliagkas
11:40 am - 12:20 am	Invited talk: Security, Privacy, and Democratization: Challenges & Future Directions for ML Systems beyond Scalability ( <a href="https://youtu.be/8YB4-HUC3n4">https://youtu.be/8YB4-HUC3n4</a> ), Dawn Song
12:20 pm - 1:30 pm	Lunch
<b>Session III</b> (moderator: Sarah Bird)	
1:30 pm - 2:10 pm	Invited talk: Structured ML: Opportunities and Challenges for the SysML Community ( <a href="https://youtu.be/7k-Ta024vQs">https://youtu.be/7k-Ta024vQs</a> ), Lise Getoor
2:10 pm - 2:20 pm	Contributed talk: Understanding the Limitations of Current Energy-Efficient Design Approaches for Deep Neural Networks ( <a href="https://youtu.be/XCdy5egmvaU">https://youtu.be/XCdy5egmvaU</a> ), Vivienne Sze
2:20 am - 2:30 am	Contributed talk: Towards High-Performance Prediction Serving Systems ( <a href="https://youtu.be/dajDalUxBsM">https://youtu.be/dajDalUxBsM</a> ), Matteo Interlandi
2:30 pm - 2:55 pm	Coffee Break
<b>Session IV</b> (moderator: Sarah Bird)	
2:55 pm - 3:05 pm	Contributed talk: "I Like the Way You Think!" - Inspecting the Internal Logic of Recurrent Neural Networks ( <a href="https://youtu.be/ErLUBbspNRY">https://youtu.be/ErLUBbspNRY</a> ), Thibault Sellam
3:05 pm - 3:45 pm	Invited talk: Systems and Machine Learning Symbiosis ( <a href="https://youtu.be/Nj6uxDki6-0">https://youtu.be/Nj6uxDki6-0</a> ), Jeff Dean
3:45 pm - 4:00 pm	Closing Remarks: Matei Zaharia
<b>Poster Session, Location: Tresidder Oak Lounge</b>	
4:30 pm - 6:00 pm	Poster Session I
6:00 pm - 7:30 pm	Poster Session II

TOP

## Location



**Conference Location:** Paul Brest Hall ([https://www.google.com/maps?q=paul+brest+hall+stanford&rlz=1C5CHFA\\_enUS717US718&um=1&ie=UTF-8&sa=X&ved=0ahUKEwjeyd3Q96DZAhXizVQKHSpODmYQ\\_AUICigB](https://www.google.com/maps?q=paul+brest+hall+stanford&rlz=1C5CHFA_enUS717US718&um=1&ie=UTF-8&sa=X&ved=0ahUKEwjeyd3Q96DZAhXizVQKHSpODmYQ_AUICigB))

**Poster Session Location:** Tresidder Union (<https://www.google.com/maps/place/Tresidder+Memorial+Union,+459+Lagunita+Dr,+Stanford,+CA+94305/@37.4243048,-122.1715932,17z/data=!4m5!3m4!1s0x808fbad4451d1567:0x8820f24eeb7fde1e!8m2!3d37.4243048!4d-122.1708422>)  
*Stanford University*

## Directions & Parking

### Directions

Directions to Paul Brest Hall ([doc/directions.pdf](#))

### Parking

The closest parking lot is Wilbur Field Garage (<https://www.google.com/maps/place/Wilbur+Field+Garage/@37.4229423,-122.1657043,17z/data=!3m1!4b1!4m5!3m4!1s0x808fbad0833ff2d9:0x18240c423c4385b0!8m2!3d37.4229423!4d-122.1635156>) (Parking Structure 6).

TOP

Parking Instructions ([doc/parking.pdf](#))

## Hotels

### Room Blocks

~~Cardinal Hotel~~ ([https://res.windsurfercrs.com/bbe/page2.aspx?pcode=GEM\\_581&propertyid=14212&nights=1&checkin=02/14/2018&group=Sysml](https://res.windsurfercrs.com/bbe/page2.aspx?pcode=GEM_581&propertyid=14212&nights=1&checkin=02/14/2018&group=Sysml))

The room block is now closed. Please check for availability directly on the hotel website (<http://www.cardinalhotel.com/>).

### Other Suggestions

Stanford Guest House (<https://rde.stanford.edu/hospitality/stanford-guest-house>)

Red Cottage Inn (<http://www.redcottageinn.com/#gref>)

Stanford's Lodging Guide (<https://visit.stanford.edu/plan/lodging.html>)

## Steering Committee

Jennifer Chayes

Bill Dally

Jeff Dean

Michael I. Jordan

Yann LeCun

Fei-Fei Li

Alex Smola

Dawn Song

Eric Xing

## Organizing Committee

Peter Bailis

Sarah Bird

Dimitris Papailiopoulos (Program Co-Chair)

Chris Ré

Ben Recht

Virginia Smith

Ameet Talwalkar (Program Chair)

Matei Zaharia

[TOP](#)

## Program Committee

David Andersen, Bryan Catanzaro, Eric Chung, Christopher De Sa, Inderjit Dhillon, Alex Dimakis, Charles Elkan, Greg Ganger, Lise Getoor, Phillip Gibbons, Garth Gibson, Joseph Gonzalez, Furong Huang, Kevin Jamieson, Yangqing Jia, Rania Khalaf, Jason Knight, Tim Kraska, Aparna Lakshmiratan, Samuel Madden, Brendan McMahan, Ioannis Mitliagkas, Rajat Monga, Derek Murray, Kunle Olukotun, Theodoros Rekatsinas, Afshin Rostamizadeh, Siddhartha Sen, Evan Sparks, Ion Stoica, Shivaram Venkataraman, Rashmi Vinayak, Markus Weimer, Ce Zhang

## Submission

Authors are encouraged to submit ongoing or recent research at the intersection of systems and machine learning. The SysML Program Committee will consider all submitted papers and will decide which papers are to be presented at the conference as oral or poster presentations.

We are interested in areas including, but not limited to, the following topics:

- Distributed and parallel learning algorithms
- Hardware-efficient model architectures
- Efficient model inference / model serving
- Visualization of data, models, and predictions

[TOP](#)

- Fairness and interpretability for ML applications
- Privacy and security for ML applications
- Customized hardware
- ETL, data preparation, feature selection, feature extraction
- Testing, debugging, and monitoring of ML applications
- ML system interfaces
- ML programming models and abstractions

**Note:** *This year, accepted submissions will be considered non-archival and can be submitted elsewhere without modification. Moreover, submissions to SysML based on recently published work are also acceptable (though authors should explicitly make note of this in their submissions). After the inaugural conference, SysML will transition to a full conference with long-form paper submissions and proceedings, to run in early 2019.*

**Submit papers via CMT:** <https://cmt3.research.microsoft.com/SYSML2018> (<https://cmt3.research.microsoft.com/SYSML2018>)

**Submission format:** Submitted abstracts should be up to 2 pages long (not including references). Submissions are not blind: author names and affiliations should appear on the first page. Papers should be formatted using the 2017 ACM Master Article Template (<https://www.acm.org/publications/proceedings-template>). For LaTeX users, choose `\format=sigconf`. This is the typical, two-column proceedings-style template. Authors do not need to include terms, keywords, or other front matter in their submissions.

**Submission deadline:** January 5th, 2018 at 11:59pm PST.

**Notification of acceptance:** ~~January 12th, 2018~~ January 16th, 2018.

Tweet

TOP

# Sponsors

*Thank you to our 2018 sponsors for your generous support!*

## Platinum Sponsors



facebook

Google

Microsoft

## General Sponsors

Amplify  
VENTURES

aws

DEFENSE INNOVATION  
UNIT EXPERIMENTAL

Determined AI

IBM Research AI

intel

Matroid

NETFLIX

TWO SIGMA

UBER  
Engineering

vertex  
VENTURES

# Contact

Please use the following email to inquire about conference information:

[TOP](#)



[info@sysml.cc](mailto:info@sysml.cc) (<mailto:info@sysml.cc>)

## Accepted Posters

Note to authors: All posters should be 3 feet wide x 4 feet tall max.

### Session I: 4:30pm – 6:00pm

1-1	A SIMD-MIMD Acceleration with Access-Execute Decoupling for Generative Adversarial Networks (doc/4.pdf)	Amir Yazdanbakhsh, Kambiz Samadi, Hadi Esmaeilzadeh, Nam Sung Kim
1-2	Slice Finder: Automated Data Slicing for Model Interpretability (doc/8.pdf)	Yeounoh Chung, Tim Kraska, Steven Euijong Whang, Neoklis Polyzotis
1-3	Data Infrastructure for Machine Learning (doc/9.pdf)	Eric Breck, Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, Martin Zinkevich
1-4	Speeding up ImageNet Training on Supercomputers (doc/10.pdf)	Yang You, Zhao Zhang, Cho-Jui Hsieh, James Demmel, Kurt Keutzer
1-5	Aloha: A Machine Learning Framework for Engineers (doc/13.pdf)	Ryan M Deak, Jonathan H Morra
1-6	Parameter Hub: High Performance Parameter Servers for Efficient Distributed Deep Neural Network Training (doc/16.pdf)	Liang Luo, Jacob Nelson, Luis Ceze, Amar Phanishayee, Arvind Krishnamurthy
1-7	Stitch-X: An Accelerator Architecture for Exploiting Unstructured Sparsity in Deep Neural Networks (doc/17.pdf)	Ching-En Lee, Yakun Sophia Shao, Jie-Fang Zhang, Angshuman Parashar, Joel Emer, Stephen W. Keckler, Zhengya Zhang

[TOP](#)

1-8	DeepVizdom: Deep Interactive Data Exploration (doc/19.pdf)	Carsten Binnig, Kristian Kersting, Alejandro Molina, Emanuel Zraggen
1-9	Massively Parallel Video Networks (doc/20.pdf)	João Carreira, Viorica Pătrăucean, Andrew Zisserman, Simon Osindero
1-10	EVA: An Efficient System for Exploratory Video Analysis (doc/22.pdf)	Ziqiang Feng, Junjue Wang, Jan Harkes, Padmanabhan Pillai, Mahadev Satyanarayanan
1-11	Declarative Metadata Management: A Missing Piece in End-To-End Machine Learning (doc/23.pdf)	Sebastian Schelter, Joos-Hendrik Böse, Johannes Kirschnick, Thoralf Klein, Stephan Seufert
1-12	Runway: machine learning model experiment management tool (doc/26.pdf)	Jason Tsay, Todd Mummert, Norman Bobroff, Alan Braz, Peter Westerink, Martin Hirzel
1-13	STRADS-AP: Simplifying Distributed Machine Learning Programming (doc/27.pdf)	Jin Kyu Kim, Garth A. Gibson, Eric P. Xing
1-14	A Deeper Look at FFT and Winograd Convolutions (doc/28.pdf)	Aleksandar Zlateski, Zhen Jia, Kai Li, Fredo Durand
1-15	Efficient Deep Learning Inference on Edge Devices (doc/29.pdf)	Ziheng Jiang, Tianqi Chen, Mu Li
1-16	On Human Intellect and Machine Failures: Troubleshooting Integrative Machine Learning Systems (doc/98.pdf)	Besmira Nushi, Ece Kamar, Eric Horvitz, Donald Kossmann
1-17	DeepThin: A Self-Compressing Library for Deep Neural Networks (doc/33.pdf)	Matthew Sotoudeh, Sara S. Bagsorkhi
1-18	MAERI: Enabling Flexible Dataflow Mapping over DNN Accelerators via Programmable Interconnects (doc/34.pdf)	Hyoukjun Kwon, Ananda Samajdar, Tushar Krishna
1-19	On Machine Learning and Programming Languages (doc/37.pdf)	Mike Innes, Stefan Karpinski, Viral Shah, David Barber, Pontus Stenetorp, Tim Besard, James Bradbury, Valentin Churavy, Simon Danisch, Alan Edelman, Jon Malmaud, Jarrett Revels, Deniz Yuret
1-20	"I Like the Way You Think!" - Inspecting the Internal Logic of Recurrent Neural Networks (doc/38.pdf)	Thibault Sellam, Kevin Lin, Ian Yiran Huang, Carl Vondrick, Eugene Wu
1-21	Automatic Differentiation in Myia (doc/39.pdf)	Olivier Breuleux, Bart van Merriënboer
1-22	TFX Frontend: A Graphical User Interface for a Production-Scale Machine Learning Platform (doc/40.pdf)	Peter Brandt, Josh Cai, Tommie Gannert, Pushkar Joshi, Rohan Khot, Chiu Yuen Koo, Chenkai Kuang, Sammy Leong, Clemens Mewald, Neoklis Polyzotis, Herve Quiroz, Sudip Roy, Po-Feng Yang, James Wexler, Steven Euijong Whang
1-23	Learned Index Structures (doc/43.pdf)	Tim Kraska, Alex Beutel, Ed H. Chi, Jeffrey Dean, Neoklis Polyzotis

TOP

1-24	Towards Optimal Winograd Convolution on Manycores (doc/47.pdf)	Zhen Jia, Aleksandar Zlateski, Fredo Durand, Kai Li
1-25	Mobile Machine Learning Hardware at ARM: A Systems-on-Chip (SoC) Perspective (doc/48.pdf)	Yuhao Zhu, Matthew Mattina, Paul Whatmough
1-26	Deep Learning with Apache SystemML (doc/49.pdf)	Niketan Pansare, Michael Dusenberry, Nakul Jindal, Matthias Boehm, Berthold Reinwald, Prithviraj Sen
1-27	Scalable Language Modeling: WikiText-103 on a Single GPU in 12 hours (doc/50.pdf)	Stephen Merity, Nitish Shirish Keskar, James Bradbury, Richard Socher
1-28	PipeDream: Pipeline Parallelism for DNN Training (doc/51.pdf)	Aaron Harlap, Deepak Narayanan, Amar Phanishayee, Vivek Seshadri, Gregory R. Ganger, Phillip B. Gibbons
1-29	Efficient Mergeable Quantile Sketches using Moments (doc/52.pdf)	Edward Gan, Jialin Ding, Peter Bailis
1-30	Systems Optimizations for Learning Certifiably Optimal Rule Lists (doc/54.pdf)	Nicholas Larus-Stone, Elaine Angelino, Daniel Alabi, Margo Seltzer, Vassilios Kaxiras, Aditya Saligrama, Cynthia Rudin
1-31	Accelerating Model Search with Model Batching (doc/55.pdf)	Deepak Narayanan, Keshav Santhanam, Matei Zaharia
1-32	Programming Language Support for Natural Language Interaction (doc/56.pdf)	Alex Renda, Harrison Goldstein, Sarah Bird, Chris Quirk, Adrian Sampson
1-33	Factorized Deep Retrieval and Distributed TensorFlow Serving (doc/57.pdf)	Xinyang Yi, Yi-Fan Chen, Sukriti Ramesh, Vinu Rajashekhar, Lichan Hong, Noah Fiedel, Nandini Seshadri, Lukasz Heldt, Xiang Wu, Ed H. Chi
1-34	Relaxed Pruning: Memory-Efficient LSTM Inference Engine by Limiting the Synaptic Connection Patterns (doc/58.pdf)	Jaeha Kung, Junki Park, Jae-Joon Kim
1-35	Deploying Deep Ranking Models for Search Verticals (doc/59.pdf)	Rohan Ramanath, Gungor Polatkan, Liqin Xu, Harold Lee, Bo Hu, Shan Zhou
1-36	Understanding the Error Structure as a Key to Regularize Convolutional Neural Networks (doc/62.pdf)	Bilal Alsallakh, Amin Jourabloo, Mao Ye, Xiaoming Liu, Liu Ren
1-37	On Scale-out Deep Learning Training for Cloud and HPC (doc/64.pdf)	Srinivas Sridharan, Karthikeyan Vaidyanathan, Dhiraj Kalamkar, Dipankar Das, Mikhail E. Smorkalov, Mikhail Shiryayev, Dheevatsa Mudigere, Naveen Mellempudi, Sasikanth Avancha, Bharat Kaul, Pradeep Dubey
1-38	In-network Neural Networks (doc/67.pdf)	Giuseppe Siracusano, Roberto Bifulco
1-39	Compressing Deep Neural Networks with Probabilistic Data Structures (doc/68.pdf)	Brandon Reagen, Udit Gupta, Robert Adolf, Michael M. Mitzenmacher, Alexander M. Rush, Gu-Yeon Wei, David Brooks
1-40	Greenhouse: A Zero-Positive Machine Learning System for Time-Series Anomaly Detection (doc/69.pdf)	Tae Jun Lee, Justin Gottschlich, Nesime Tatbul, Eric Metcalf, Stan Zdonik

TOP

1-41	Precision and Recall for Range-Based Anomaly Detection (doc/70.pdf)	Tae Jun Lee, Justin Gottschlich, Nesime Tatbul, Eric Metcalf, Stan Zdonik
1-42	Whetstone: An accessible, platform-independent method for training spiking deep neural networks for neuromorphic processors (doc/71.pdf)	William M. Severa, Craig M. Vineyard, Ryan Dellana, James B. Aimone
1-43	SparseCore: An Accelerator for Structurally Sparse CNNs (doc/72.pdf)	Sharad Chole, Ramteja Tadishetti, Sree Reddy
1-44	SGD on Random Mixtures: Private Machine Learning under Data Breach Threats (doc/73.pdf)	Kangwook Lee, Kyungmin Lee, Hoon Kim, Changho Suh, Kannan Ramchandran
1-45	Towards High-Performance Prediction Serving Systems (doc/74.pdf)	Yunseong Lee, Alberto Scolari, Matteo Interlandi, Markus Weimer, Byung-Gon Chun
1-46	Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization (doc/75.pdf)	Fabian Pedregosa, Rémi Leblond, Simon Lacoste-Julien
1-47	Corpus Conversion Service: A machine learning platform to ingest documents at scale. (doc/76.pdf)	Peter W J Staar, Michele Dolfi, Christoph Auer, Costas Bekas
1-48	Representation Learning for Resource Usage Prediction (doc/77.pdf)	Florian Schmidt, Mathias Niepert, Felipe Huici
1-49	TVM: End-to-End Compilation Stack for Deep Learning (doc/78.pdf)	Tianqi Chen, Thierry Moreau, Ziheng Jiang, Haichen Shen, Eddie Yan, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy
1-50	vectorflow: a minimalist neural-network library (doc/79.pdf)	Benoît Rostykus, Yves Raimond
1-51	Learning Heterogeneous Cloud Storage Configuration for Data Analytics (doc/82.pdf)	Ana Klimovic, Heiner Litz, Christos Kozyrakis
1-52	Salus: Fine-Grained GPU Sharing Among CNN Applications (doc/83.pdf)	Peifeng Yu, Mosharaf Chowdhury
1-53	OpenCL Acceleration for TensorFlow (doc/84.pdf)	Mehdi Goli, Luke Iwanski, John Lawson, Uwe Dolinsky, Andrew Richards
1-54	Picking Interesting Frames in Streaming Video (doc/85.pdf)	Christopher Canel, Thomas Kim, Giulio Zhou, Conglong Li, Hyeontaek Lim, David G. Andersen, Michael Kaminsky, Subramanya R. Dulloor
1-55	SLAQ: Quality-Driven Scheduling for Distributed Machine Learning (doc/86.pdf)	Haoyu Zhang, Logan Stafman, Andrew Or, Michael J. Freedman
1-56	A Comparison of Bottom-Up Approaches to Grounding for Templated Markov Random Fields (doc/87.pdf)	Eriq Augustine, Lise Getoor

TOP

1-57	Growing Cache Friendly Decision Trees (doc/89.pdf)	Niloy Gupta, Adam Johnston
1-58	Parallelizing Hyperband for Large-Scale Tuning (doc/100.pdf)	Lisha Li, Kevin Jamieson, Afshin Rostamizadeh, Ameet Talwalkar
1-59	Towards Interactive Curation and Automatic Tuning of ML Pipelines (doc/118.pdf)	Carsten Binnig, Benedetto Buratti, Yeounoh Chung, Cyrus Cousins, Dylan Ebert, Tim Kraska, Zeyuan Shang, Isabella Tromba, Eli Upfal, Linnan Wang, Robert Zeleznik, Emanuel Zraggen

## Session II: 6:00pm – 7:30pm

2-1	Ternary Residual Networks (doc/24.pdf)	Abhisek Kundu, Kunal Banerjee, Naveen Mellempudi, Dheevatsa Mudigere, Dipankar Das, Bharat Kaul, Pradeep Dubey
2-2	Neural Architect: A Multi-objective Neural Architecture Search with Performance Prediction (doc/94.pdf)	Yanqi Zhou, Gregory Diamos
2-3	Federated Kernelized Multi-Task Learning (doc/30.pdf)	Sebastian Caldas, Virginia Smith, Ameet Talwalkar
2-4	Materialization Trade-offs for Feature Transfer from Deep CNNs for Multimodal Data Analytics (doc/101.pdf)	Supun Nakandala, Arun Kumar
2-5	Scaling HDBSCAN Clustering with kNN Graph Approximation (doc/105.pdf)	Jacob Jackson, Aurick Qiao, Eric P. Xing
2-6	Blazelt: An Optimizing Query Engine for Video at Scale (doc/107.pdf)	Daniel Kang, Peter Bailis, Matei Zaharia
2-7	Time Travel based Feature Generation (doc/108.pdf)	Kedar Sadekar, Hua Jiang
2-8	Controlling AI Engines in Dynamic Environments (doc/109.pdf)	Nikita Mishra, Connor Imes, Henry Hoffmann, John D. Lafferty
2-9	Intermittent Deep Neural Network Inference (doc/110.pdf)	Graham Gobieski, Nathan Beckmann, Brandon Lucia
2-10	CascadeCNN: Pushing the performance limits of quantisation (doc/114.pdf)	Alexandros Kouris, Stylianos I. Venieris, Christos-Savvas Bouganis
2-11	Making Machine Learning Easy with Embeddings (doc/115.pdf)	Dan Shiebler, Abhishek Tayal
2-12	CrossBow: Scaling Deep Learning on Multi-GPU Servers (doc/119.pdf)	Alexandros Koliousis, Pijika Watcharapichat, Matthias Weidlich, Paolo Costa, Peter Pietzuch

TOP

2-13	Better Caching with Machine Learned Advice (doc/121.pdf)	Thodoris Lykouris, Sergei Vassilvitskii
2-14	Large Model Support for Deep Learning in Caffe and Chainer (doc/127.pdf)	Minsik Cho, Tung D. Le, Ulrich A. Finkler, Haruiki Imai, Yasushi Negishi, Taro Sekiyama, Saritha Vinod, Vladimir Zolotov, Kiyokuni Kawachiya, David S. Kung, Hillery C. Hunter
2-15	Learning Graph-based Cluster Scheduling Algorithms (doc/129.pdf)	Hongzi Mao, Malte Schwarzkopf, Shaileshh Bojja Venkatakrishnan, Mohammad Alizadeh
2-16	Intel nGraph: An Intermediate Representation, Compiler, and Executor for Deep Learning (doc/132.pdf)	Scott Cyphers, Arjun K. Bansal, Anahita Bhiwandiwalla, Jayaram Bobba, Matthew Brookhart, Avijit Chakraborty, Will Constable, Christian Convey, Leona Cook, Omar Kanawi, Robert Kimball, Jason Knight, Nikolay Korovaiko, Varun Kumar, Yixing Lao, Christopher R. Lishka, Jaikrishnan Menon, Jennifer Myers, Sandeep Aswath Narayana, Adam Procter, Tristan J. Webb
2-17	Efficient Multi-Tenant Inference on Video using Microclassifiers (doc/133.pdf)	Giulio Zhou, Thomas Kim, Christopher Canel, Conglong Li, Hyeontaek Lim, David G. Andersen, Michael Kaminsky, Subramanya R. Dulloor
2-18	Abstractions for Containerized Machine Learning Workloads in the Cloud (doc/135.pdf)	Balaji Subramaniam, Niklas Nielsen, Connor Doyle, Ajay Deshpande, Jason Knight, Scott Leishman
2-19	Not All Ops Are Created Equal! (doc/137.pdf)	Liangzhen Lai, Naveen Suda, Vikas Chandra
2-20	Robust Gradient Descent via Moment Encoding with LDPC Codes (doc/140.pdf)	Raj Kumar Maity, Ankit Singh Rawat, Arya Mazumdar
2-21	Buzzsaw: A System for High Speed Feature Engineering (doc/141.pdf)	Andrew Stanton, Liangjie Hong, Manju Rajashekhar
2-22	Predicate Optimization for a Visual Analytics Database (doc/143.pdf)	Michael R. Anderson, Michael Cafarella, Thomas F. Wenisch, German Ros
2-23	Understanding the Limitations of Current Energy-Efficient Design Approaches for Deep Neural Networks (doc/145.pdf)	Yu-Hsin Chen, Tien-Ju Yang, Joel Emer, Vivienne Sze
2-24	Compiling machine learning programs via high-level tracing (doc/146.pdf)	Roy Frostig, Matthew James Johnson, Chris Leary
2-25	Dynamic Stem-Sharing for Multi-Tenant Video Processing (doc/149.pdf)	Angela Jiang, Christopher Canel, Daniel Wong, Michael Kaminsky, Michael A. Kozuch, Padmanabhan Pillai, David G. Andersen, Gregory R. Ganger
2-26	A Hierarchical Model for Device Placement (doc/150.pdf)	Azalia Mirhoseini, Anna Goldie, Hieu Pham, Benoit Steiner, Quoc V. Le, Jeff Dean

TOP

2-27	Blink: A fast NVLink-based collective communication library (doc/151.pdf)	Guanhua Wang, Amar Phanishayee, Shivaram Venkataraman, Ion Stoica
2-28	TOP: A Compiler-Based Framework for Optimizing Machine Learning Algorithms through Generalized Triangle Inequality (doc/152.pdf)	Yufei Ding, Lin Ning, Hui Guang, Xipeng Shen, Madanlal Musuvathi, Todd Mytkowicz
2-29	UberShuffle: Communication-efficient Data Shuffling for SGD via Coding Theory (doc/157.pdf)	Jichan Chung, Kangwook Lee, Ramtin Pedarsani, Dimitris Papailiopoulos, Kannan Ramchandran
2-30	Toward Scalable Verification for Safety-Critical Deep Networks (doc/158.pdf)	Lindsey Kuper, Guy Katz, Justin Gottschlich, Kyle Julian, Clark Barrett, Mykel J. Kochenderfer
2-31	DAWNBench: An End-to-End Deep Learning Benchmark and Competition (doc/159.pdf)	Cody Coleman, Deepak Narayanan, Daniel Kang, Tian Zhao, Jian Zhang, Luigi Nardi, Peter Bailis, Kunle Olukotun, Chris Ré, Matei Zaharia
2-32	Learning Network Size While Training with ShrinkNets (doc/161.pdf)	Guillaume Leclerc, Raul Castro Fernandez, Samuel Madden
2-33	Have a Larger Cake and Eat It Faster Too: A Guideline to Train Larger Models Faster (doc/162.pdf)	Newsha Ardalani, Joel Hestness, Gregory Diamos
2-34	Retrieval as a defense mechanism against adversarial examples in convolutional neural networks (doc/163.pdf)	Junbo Zhao, Jinyang Li, Kyunghyun Cho
2-35	DNN-Train: Benchmarking and Analyzing Deep Neural Network Training (doc/167.pdf)	Hongyu Zhu, Bojian Zheng, Bianca Schroeder, Gennady Pekhimenko, Amar Phanishayee
2-36	High Accuracy SGD Using Low-Precision Arithmetic and Variance Reduction (for Linear Models) (doc/169.pdf)	Alana Marzoev, Christopher De Sa
2-37	SkipNet: Learning Dynamic Routing in Convolutional Networks (doc/170.pdf)	Xin Wang, Fisher Yu, Zi-Yi Dou, Joseph E. Gonzalez
2-38	Memory-Efficient Data Structures for Learning and Prediction (doc/171.pdf)	Damian Eads, Paul Baines, Joshua S. Bloom

TOP

2-39	Efficient and Programmable Machine Learning on Distributed Shared Memory via Static Analysis (doc/172.pdf)	Jinliang Wei, Garth A. Gibson, Eric P. Xing
2-40	Parle: parallelizing stochastic gradient descent (doc/174.pdf)	Pratik Chaudhari, Carlo Baldassi, Riccardo Zecchina, Stefano Soatto, Ameet Talwalkar, Adam Oberman
2-41	Optimal Message Scheduling for Aggregation (doc/178.pdf)	Leyuan Wang, Mu Li, Edo Liberty, Alex J. Smola
2-42	Analog electronic deep networks for fast and efficient inference (doc/179.pdf)	Jonathan Binas, Daniel Neil, Giacomo Indiveri, Shih-Chii Liu, Michael Pfeiffer
2-43	Network Evolution for DNNs (doc/182.pdf)	Michael Alan Chang, Aurojit Panda, Domenic Bottini, Lisa Jian, Pranay Kumar, Scott Shenker
2-44	BinaryCmd: Keyword Spotting with deterministic binary basis (doc/183.pdf)	Javier Fernández-Marqués, Vincent W.-S. Tseng, Sourav Bhattachara, Nicholas D. Lane
2-45	YellowFin: Adaptive Optimization for (A)synchronous Systems (doc/188.pdf)	Jian Zhang, Ioannis Mitliagkas
2-46	GPU-acceleration for Large-scale Tree Boosting (doc/192.pdf)	Huan Zhang, Si Si, Cho-Jui Hsieh
2-47	Treelite: toolbox for decision tree deployment (doc/196.pdf)	Hyunsu Cho, Mu Li
2-48	On Importance of Execution Ordering in Graph-Based Distributed Machine Learning Systems (doc/197.pdf)	Sayed Hadi Hashemi, Sangeetha Abdu Jyothi, Roy Campbell
2-49	Draco: Robust Distributed Training against Adversaries (doc/199.pdf)	Lingjiao Chen, Hongyi Wang, Dimitris Papailiopoulos
2-50	Clustering System Data using Aggregate Measures (doc/201.pdf)	Johnnie C-N. Chang, Robert H-J. Chen, Jay Pujara, Lise Getoor
2-51	A Framework for Searching a Predictive Model (doc/202.pdf)	Yoshiki Takahashi, Masato Asahara, Kazuyuki Shudo
2-52	Distributed Placement of Machine Learning Operators for IoT applications spanning Edge and Cloud Resources (doc/204.pdf)	Tarek Elgamal, Atul Sandur, Klara Nahrstedt, Gul Agha
2-53	Finding Heavily-Weighted Features with the Weight-Median Sketch (doc/205.pdf)	Kai Sheng Tai, Vatsal Sharan, Peter Bailis, Gregory Valiant

TOP



2-54	Flexible Primitives for Distributed Deep Learning in Ray (doc/206.pdf)	Yaroslav Bulatov, Robert Nishihara, Philipp Moritz, Melih Elibol, Ion Stoica, Michael I. Jordan
2-55	BLAS-on-flash: an alternative for training large ML models? (doc/207.pdf)	Suhas Jayaram Subramanya, Srajan Garg, Harsha Vardhan Simhadri
2-56	Treating Machine Learning Algorithms As Declaratively Specified Circuits (doc/210.pdf)	Jason Eisner, Nathaniel Wesley Filardo
2-57	Tasvir: Distributed Shared Memory for Machine Learning (doc/214.pdf)	Amin Tootoonchian, Aurojit Panda, Aida Nematzadeh, Scott Shenker

## Conference Code of Conduct

All attendees, speakers, sponsors and volunteers at our conference are required to agree with the following code of conduct. Organisers will enforce this code throughout the event. We expect cooperation from all participants to help ensure a safe environment for everybody.

*Our conference is dedicated to providing a harassment-free conference experience for everyone, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), or technology choices. We do not tolerate harassment of conference participants in any form. Sexual language and imagery is not appropriate for any conference venue, including talks, workshops,*

TOP

*parties, Twitter and other online media. Conference participants violating these rules may be sanctioned or expelled from the conference without a refund at the discretion of the conference organizers.*

For more information, visit: <http://confcodeofconduct.com/>  
(<http://confcodeofconduct.com/>)

TOP