Paperscope: Chronicling the History of Computer Architecture Research

Vincent T. Lee, Luis Ceze, Mark Oskin
University of Washington
{vlee2, luisceze, oskin}@cs.washington.edu

***Abstract*—Computer architecture, like many other well established research fields, has witnessed periodic trends in research interests. Certain research topics transcend the test of time, while others enjoy surges in popularity followed by a period of decline and interest. However, there is little actual data that corroborates or rebukes these patterns of popularity within our community. To substantiate these observations, we use latent Dirichlet allocation (LDA) to build a topic model over all publications from three top tier computer architecture conferences up until 2015. Our results show that LDA provides an excellent method for chronicling the trajectory of the computer architecture research community, and accurately quantifies historical trends and patterns in computer architecture research.**

***Index Terms*—latent Dirichlet allocation, history, topic model**

I. Introduction

Using latent Dirichlet allocation (LDA) [2] to build topic models has been shown to be extremely effective at identifying similar content and for knowledge discovery across a corpus of documents in computational linguistics [3], social media [6], and even source code histories [9]. LDA is an unsupervised learning algorithm which attempts to derive sets of keywords or *vocabularies* which occur frequently together to form *topics*. For instance, the topic vocabulary for the topic “branch prediction” may be composed of the words “branch”, “predictor”, “mispredict”, “prediction”, “accuracy”, and “control”. Each keyword is also ascribed a weight which indicates how strongly it identifies with the topic. In the previous example, we may ascribe higher weight to the keywords “branch” and “predictor” since they are good indicators that the document is about branch prediction, and ascribe lower weight to the word “control” since a paper which contains many instances of the word “control” does not necessarily mean it’s about branch prediction. Keywords in a vocabulary are not unique to a single topic and can appear across multiple topics with different weight values in the model.

Together, topic vocabularies and weights form a *topic model*, which can be used to mine common patterns among an existing corpus of documents, extract document features, or determine document similarity [11]. In our work we train a topic model across all available documents published across three of the top tier computer architectures conferences up until 2015: International Symposium on Computer Architecture (ISCA), International Symposium on Microarchitecture (MICRO), and International Conference on Architecture Systems Support for Programming Languages and Operating Systems (ASPLOS). We then chronicle the resulting topics by year to characterize the high level research trajectories and explore the historical trends of research topics across our field.

In particular we are interested in exploring how research focuses and goals of the computer architecture community have changed, diverged, and evolved over time. Such insight can be invaluable to the community when auditing related work, evaluating renewed relevance of past ideas, and identifying the overall trajectory of the research community over time. This study aims to answer the following questions: (1) What shifts in computer architecture research have happened in the past? (2) Is the research in our community becoming more broad or more narrow? and (3) What research topics are emerging and which are no longer trending?

II. Methodology

For our document corpus, we use all of the available proceedings from the ACM Digital Library up to the end of 2015 for ISCA, MICRO, and ASPLOS which are widely acknowledged as the oldest top tier computer architecture conference venues in our field. We omit several proceedings that are unavailable on the ACM digital library such as the MICRO 1-5 and MICRO 7 proceedings. We also do not include workshop proceedings with the exception of MICRO which used to be held as the Annual Workshop on Microprogramming. In total, our document corpus contains over 3700 PDF documents spanning from 1972 to 2015.

To recover text content from the PDF documents, we use the Tesseract Optical Character Recognition Library [8] by first converting the document to a 300 DPI image. In certain cases Tesseract mangles adjacent letters or splits words; however, we do not expect this to affect the topic model which is resilient to this noise. After converting the documents to raw text, we use the Mallet tool [7] to generate a topic model over the corpus using 200 topics. For each document, the Mallet tool builds a feature vector where the Nth dimension corresponds to the “fraction” of the document that is about the Nth topic formed in the model. For instance, if a document is scored such that topic A gets a weight or *strength* of 0.25, and topic B gets 0.75. We can conclude that 25% of the document is about topic A, and 75% of the document is about topic B, and write a *feature vector* for this document as [0.25, 0.75].

After training the topic model, we the use the methodology in [3] and bucket the documents by year of publication. We then generate a time series for each topic which plots the aggregate strength of each topic over time; one time series is generated for each topic in the model. The aggregate topic strength for a given topic T for a given year Y is the sum of the feature vector weights for T over all documents published in Y. The resulting time series of aggregate topic strengths represents how many effective publications were about a given topic that year. We note that the actual number of documents which contribute the aggregate strength of the topic in a given year is actually many more. For instance, if the aggregate strength of a topic is 1.3 for a given year, four papers with strength 0.4, 0.1, 0.5, and 0.3 could have been published that year for that topic.

III. Results and Analysis

A. Topic Labeling and Validation

Since LDA is an unsupervised algorithm, we do now know apriori what the vocabularies for each topic will be. Furthermore, the algorithm also does not guarantee that topics selected are uncorrelated so documents about similar research areas may manifest across multiple topics (as we will show later). In order to contextualize and validate our results with respect to actual historical and current research trends, we manually inspect the resulting set of top 10 keywords and the highest scoring documents for each topic to determine the corresponding label.

Our results unsurprisingly show that LDA is very effective at discovering and grouping documents in the corpus by topics we are familiar with. Table I shows the top ten keywords for selected topics in the model of the 200 topics that we train. Based on the resulting keywords, we can infer that the topic model forms vocabularies for well known research areas in architecture such as approximate computing, neural networks and deep learning, branch prediction, and caches. As expected, for a topic like neural networks we see keywords like "neural", "training", "weights", and "layers". Similarly for a topic like approximate computing we observe keywords like "precision", "approximation", "quality", and "error". To validate whether the topic model correctly groups publications by research areas we manually inspect the highest scoring papers for each topic. For example, the top papers with the highest strength for the topic corresponding to approximate computing are shown below:

- Rumba: An Online Quality Management System for Approximate Computing (2015)
- Quality Programmable Vector Processors for Approximate Computing (2013)
- SAGE, Self-Tuning Approximation for Graphics Engines (2013)
- Load Value Approximation (2014)
- Monitoring and Debugging the Quality of Results (2015)
- ApproxHadoop: Bringing Approximations to MapReduce Frameworks (2015)
- Paraprox, Pattern-Based Approximation (2014)
- Neural Acceleration for General-Purpose Approximate Programs (2012)

We note that the topic model score of a paper only indicates how strongly the paper content associates with that topic since the algorithm scores documents purely by word frequency and is oblivious to citation counts. A paper may also associated with multiple topics which splits feature vectors weights among multiple topics.

B. Emerging Research Topics

By organizing our results and tracking the strength of a topic temporally, we can measure the popularity or interest in a research topic over the years. For instance, for truly novel research areas we expect trend lines to exhibit no past activity followed by a recent increase in topic strength. Figure 1 shows the strength of the topics corresponding to approximate computing, security, neural networks, and programmable accelerator fabrics. As one would expect, these trend lines exhibit strong activity in recent years confirming our research community's recent surge in interest in these areas.

C. Declining Research Topics

We also observe topics which have previously demonstrated significant interest but have since fallen out of favor in the research community. For example, Figure 2 shows the trends for topics corresponding to quantum computing, Lisp machines, and VAX / 8086 / PDP era machines which show an initial rise in topic strength followed by a plateau, and an eventual decline indicating these research areas are no longer active. Going back to our roots, the topic model confirms that our research community stemmed from core microarchitecture research such as microcoded machines, logic design, and high level languages to abstract software from hardware (Figure 3). As indicated by the trend lines, these topics are no longer as active as they were towards the beginning of architecture research but have since declined in activity. What is interesting is that the trends for core architecture topics do not fully dissipate until the mid-1990s indicating strong relevance until then.

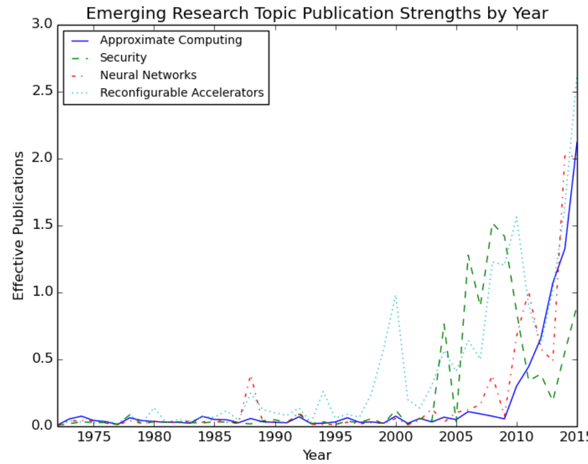

Fig. 1: Strength of currently trending research topics over time.

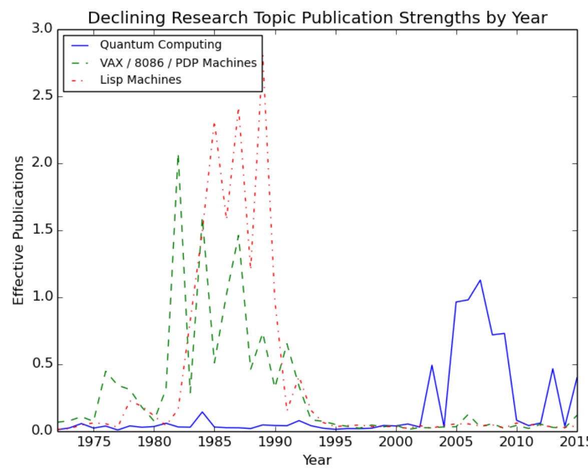

Fig. 2: Research topics that were previously popular but have experienced a decline in interest.

D. Research Withstanding the Test of Time

We find that for computer architecture research areas which withstand the test of time that the topic model splits these documents among several finer grained topics. One notable example that manifests as six topics in the model is cache research (vocabularies shown in Table II). Our results show that the topic model is able differentiate among finer granularity areas of focus in cache research such as cache coherence, and caches for chip multiprocessors (CMP). When charting the combined strength of these cache research topics, we find that individually the trends lines are unremarkable but combined they attest to the consistent strength of cache research since the 1980s (Figure 4). We observe similar splitting among topics (not shown) for research areas including branch prediction, GPU programming and architectures, security, and graph processing.

E. Our Research Community is Becoming Broader

To determine whether our research community is becoming broader as a whole, we plot the number of active topics published per year. To do this, we aggregate the strength of each topic for each year and apply a threshold value t to determine whether a topic is "active" that year. If the aggregate strength of a topic exceeds the threshold, we can conclude that the topic was active and featured in at least t

Topic Label	Keywords				
Quantum Computing	quantum correction	error gate	qubits data	qubit gates	ancilla circuit
GPU Architectures	warp thread	warps gpgpu	threads execution	divergence gpu	active figure
Cache Coherence	shared multiprocessor	cache sharing	bus processor	data invalidation	caches reference
Microcoded Architectures (μ = "micro")	micro μinstruction	control μinstructions	microcode μprograms	μprogram μprogrammed	μprogramming store
Caches	miss size	misses data	cache mapped	direct cpi	spec benchmarks
Predicated Execution	predicate branch	control predication	predicated figure	code branches	execution ow
High Level Language Design	language high	type compiler	languages level	programming code	interpreter implementation
Prefetching	prefetching miss	prefetch misses	prefetches hardware	prefetcher table	stride prefetched
Neural Networks	neural neurons	learning training	network weights	networks layer	neuron input
Security	security metadata	ow analysis	information tags	tracking software	tag hardware
VAX, 8086, PDP-10 Research	vax table	risc ll	pdp mode	microcode byte	data operand
Microarchitecture	address data	register instruction	control unit	operand alu	registers operation
Parallelism	parallel execution	parallelism ne	sequential parallelization	speedup grain	data serial
Reconfigurable Accelerators	hardware application	accelerator recon	data gurable	fpga design	accelerators custom
Approximate Computing	approximate approximation	quality application	error applications	output input	data precise

TABLE I: Top ten keywords for selected topics manifesting in topic model.

Topic Label	Keywords				
Caches	miss size	misses data	cache mapped	direct cpi	spec benchmarks
Caches	cache lines	line size	caches hit	data performance	miss access
Chip Multiprocessor (CMP) / Distributed Caches	private nuca	shared access	cmp data	chip cache	replication capacity
Cache Coherence Protocols	shared multiprocessor	cache sharing	bus processor	data invalidation	caches reference
Cache Coherence	coherence protocols	protocol requests	directory system	request cache	shared state
Set Associative Caches	tag entry	associative array	set access	tags index	address bits

TABLE II: Labels and vocabularies for all cache related topics.

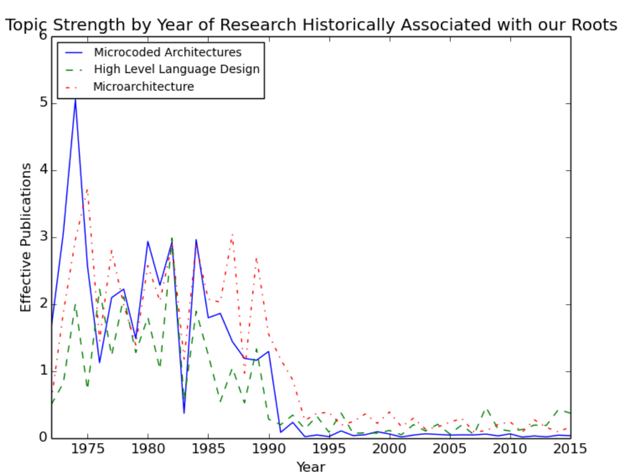

Fig. 3: Strength of research topics corresponding to those associated with the roots of architecture research.

effective publications that year. Figure 5 shows the number of topics published each year for increasing thresholds. Certain years have fewer publications (Figure 6) than others - most notably years without ASPLOS proceedings - so the number of active topics fluctuates accordingly. However, the results generally show that the number of topics has been increasing over time indicating that our research community's interests are becoming increasingly more diverse.

IV. Discussion

Our conference publication selection is by no means a complete collection of all computer architecture work. Most notably we did not include proceedings from conferences such as International Symposium on High Performance Computing Architectures (HPCA), Supercomputing (SC), and International Parallel and Distributed Processing Symposium (IPDPS). We also do not account for journal publications which complement the conference system. Finally, we do not account for relevant computer architecture works published in other core computer science fields such as databases and graphics.

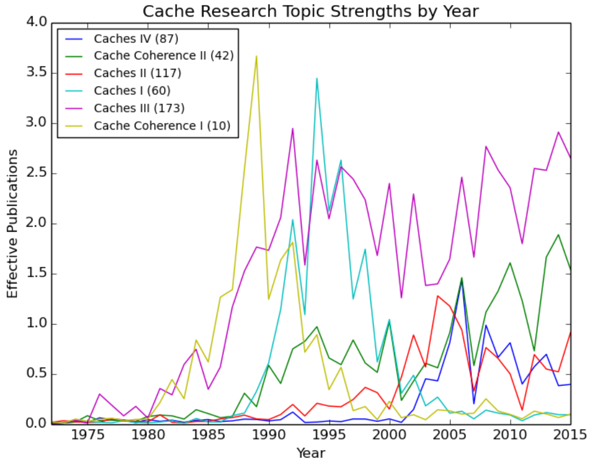

Fig. 4: Trend lines for all cache topics over time. Individually each trend line is unremarkable but combined they illustrate how resilient cache research is to the test of time.

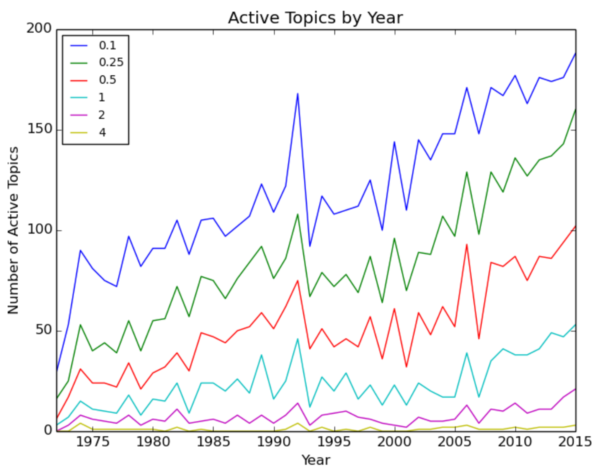

Fig. 5: Number of active topics published per year (threshold $\in \{0.1, 0.2, 0.5, 1.0, 2.0, 4.0\}$ effective publications).

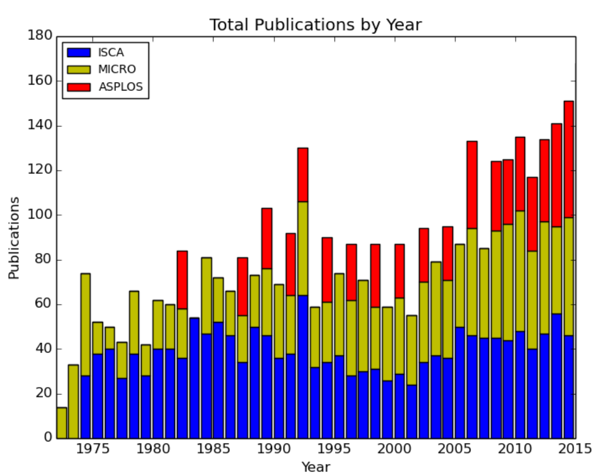

Fig. 6: Number of publications per year by conference.

V. Related Work

Applying variants of LDA to study historical trends is a well established concept in natural language processing, machine learning, and linguistics. Our approach was initially proposed by Hall et al [3] who first apply it to chronicle computational linguistics research. However other approaches such as topics over time [10] and dynamic topic models [1] have been proposed to study the evolution of a documents over time. We are also not the first to study publication trends in computer architecture, most notably Hill et al [4], [5] present a comprehensive compilation of multiprocessor research over all ISCA publications; our work augments the scope of this work further and expands it to finer grained topic areas.

VI. Conclusions

We present a study of computer architecture research history using LDA to identify historical research trends in our field. Our results show that LDA is highly effective at building a topic model to both identify active areas of research and areas which have declined in popularity. The model shows that while there are research areas which have fallen out of favor, as a whole the computer architecture community is still growing and becoming more broad in its interests. Our results also show that established research areas such as caches and branch predictions have enjoyed sustained popularity which is consistent with conference trends. Our data also corroborates the recent popularity of non-traditional areas of research such as approximate computing, security, and neural networks. Finally, our work provides quantitatively support to publication trends that may be of interest to researchers in our field.

References

[1] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: ACM, 2006, pp. 113–120. [Online]. Available: http://doi.acm.org/10.1145/1143844.1143859

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[3] D. Hall, D. Jurafsky, and C. D. Manning, "Studying the history of ideas using topic models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 363–371. [Online]. Available: http://dl.acm.org/citation.cfm?id=1613715.1613763

[4] M. D. Hill and M. R. Marty, "Amdahl's law in the multicore era," 2008.

[5] M. D. Hill and R. Rajwar, "The rise and fall of multiprocessor papers in the international symposium on computer architecture," http://pages.cs.wisc.edu/~markhill/mp2001.html, 2001, accessed: 2016-05-23.

[6] L. Hong and B. D. Davison, "Empirical study of topic modeling in twitter," in *Proceedings of the First Workshop on Social Media Analytics*, ser. SOMA '10. New York, NY, USA: ACM, 2010, pp. 80–88. [Online]. Available: http://doi.acm.org/10.1145/1964858.1964870

[7] A. K. McCallum, "Mallet: A machine learning for language toolkit," 2002, http://mallet.cs.umass.edu.

[8] R. Smith, "An overview of the tesseract ocr engine," in *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, ser. ICDAR '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 629–633. [Online]. Available: http://dl.acm.org/citation.cfm?id=1304596.1304846

[9] S. W. Thomas, B. Adams, A. E. Hassan, and D. Blostein, "Modeling the evolution of topics in source code histories," in *Proceedings of the 8th Working Conference on Mining Software Repositories*, ser. MSR '11. New York, NY, USA: ACM, 2011, pp. 173–182. [Online]. Available: http://doi.acm.org/10.1145/1985441.1985467

[10] X. Wang and A. McCallum, "Topics over time: A non-markov continuous-time model of topical trends," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '06. New York, NY, USA: ACM, 2006, pp. 424–433. [Online]. Available: http://doi.acm.org/10.1145/1150402.1150450

[11] X. Wei and W. B. Croft, "Lda-based document models for ad-hoc retrieval," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '06. New York, NY, USA: ACM, 2006, pp. 178–185. [Online]. Available: http://doi.acm.org.offcampus.lib.washington.edu/10.1145/1148170.1148204