

# Toward Metrics of Design Automation Research Impact (invited paper)

Andrew B. Kahng<sup>‡,†</sup>, Mulong Luo<sup>‡</sup>, Gi-Joon Nam<sup>1</sup>, Siddhartha Nath<sup>‡</sup>, David Z. Pan<sup>2</sup> and Gabriel Robins<sup>3</sup>

UC San Diego, <sup>†</sup>ECE and <sup>‡</sup>CSE Depts., La Jolla, CA 92093, {abk, muluo, sinath}@ucsd.edu

<sup>1</sup>IBM Research, Yorktown Heights, NY 10598, {gnam@us.ibm.com}

<sup>2</sup>UT Austin, ECE Dept., Austin, TX 78712, {dpan@ece.utexas.edu}

<sup>3</sup>Univ. of Virginia, CS Dept., Charlottesville, VA 22904, {robins@cs.virginia.edu}

**Abstract**—Design automation (DA) research has for over fifty years been performed in academia, semiconductor and system companies, and EDA companies worldwide. This research has been enabling to continued scaling of design productivity and growth of the semiconductor industry. For product companies, funding program managers and individual researchers alike, a highly relevant question is: what DA research, and what DA research outcomes, ultimately have the greatest “impact”? In this paper, we present measurements and analyses of DA research *outputs* (papers, patents, EDA companies), upon which future metrics of DA research *impact* might be based. Our studies consider 47000+ conference and journal papers from 1964-2014; the inter-patent citation graph over 759000+ DA-related patents; abstracts of 1150+ U.S. NSF projects over a three-decade span; 36 research needs documents of the Semiconductor Research Corporation from 2000-2013; and market segmentation of hundreds of EDA companies. We identify several interesting correlations, but do not claim to identify causal relationships; indeed, connecting traditional measures of research output to real-world impacts seems quite challenging. We conclude with several directions and targets for future investigation.

## I. INTRODUCTION

Design automation (DA) research has for over fifty years been performed in academia, semiconductor and system companies, and EDA companies worldwide. The genesis of DA research ranges from inspirations to zeitgeist to targeted funding streams from consortia and government agencies. The outcomes of DA research are manifested in publications, patents, commercial products, companies, and the training of future technical contributors. The dynamics of research impact are complex: research builds on previous research; a seminal research outcome can lead to a funding program which further develops the original outcome; etc. Figure 1 illustrates types of entities and research products in the EDA landscape, along with the “life cycle of ideas” seen at major conferences such as the Design Automation Conference.<sup>1</sup> Uncovering “research impact” requires understanding (identification, characterization) of the edges and feedback paths that we have omitted from the figure (e.g., edges from conference papers to journal papers, or from patents to publications; paths from consortia research needs through to papers and EDA startups; etc.).

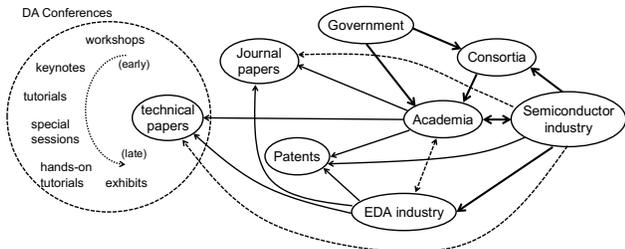


Fig. 1: Antecedents and outputs of DA research. The “life cycle” of ideas at DA conferences is also shown. Note that the question of “research impact” seeks to understand edges and feedback paths that have been omitted from this figure.

As in most scientific and engineering research arenas, “metrics” of DA research impact are difficult to formulate beyond the usual paper counts, citation counts, h-indices, i-indices, etc. (joined in recent years by “test of time” awards). At the same time, any ability to measure the impact of DA research has obvious potential benefit

<sup>1</sup>ABK attributes the conference life cycle analysis to Dr. Leon Stok of IBM, in 46th DAC (2009) executive committee discussions.

for the field, e.g., (i) as the basis for early identification of high-impact or high-value research directions and results; (ii) as part of a feedback or learning loop that helps funding agencies create higher-impact research programs with limited resources; or (iii) as motivation for increased overall investment in DA research if the ROI is sufficiently high.

Our work studies analyses of DA research *outputs* (papers, patents, EDA companies), upon which future metrics of DA research *impact* might be based. These analyses include (i) evolution over time of LDA-based topic models and ‘word clouds’, (ii) temporal offsets of topic models between various corpora (e.g., research funding vs. patents vs. papers), and (iii) topological studies of the patent citation graph.<sup>2</sup> Our example analyses consider non-trivial fractions of the history of DA research and its impacts: the text of 47000+ conference and journal papers from 1964-2014; the inter-patent citation graph over 759000+ DA-related patents; abstracts of 1150+ U.S. National Science Foundation-funded projects over a three-decade span; 36 research needs documents of the Semiconductor Research Corporation from 2000-2013; and the evolution of EDA industry segmentation over time. While analogous studies have been performed in other areas of the scientific and patent literatures [2] [23] [30], to our knowledge we are the first to perform such a comprehensive study in the DA field. Our contributions include:

- We bring together, more comprehensively than previous works of which we are aware, multiple DA research corpora: patents, conference and journal papers, government-funded research project topics, industry consortia needs statements, etc.
- We apply latent Dirichlet analysis (LDA) [7] to extract topic models of DA research corpora (patents, papers, needs statements, project abstracts) over time.
- We illustrate *alignment analyses* that suggest leading or lagging relationships among patents, papers, funding programs and industry demographics.
- We illustrate standard citation graph analyses (centrality, transitive fanout, etc.) for DA-related patents.
- We conclude with directions “toward metrics of DA research impact”, along with contemplated further analyses.

Section 2 below reviews two relevant contexts for our analyses. Section 3 describes methodologies for data collection and analyses. Section 4 presents illustrative analysis results, while Section 5 gives potential real-world “impacts” to which DA research may one day be causally tied, along with our conclusions.

## II. CONTEXT-SETTING

DA research has “monetized”, real-world impact in the EDA and semiconductor industries. We therefore begin by setting out two relevant contexts for our analyses: (i) EDA industry demographics, and (ii) semiconductor industry-provided support for DA research. **EDA industry demographics, revenues and valuations.** The EDA industry – itself an “impact” of DA research – enjoyed

<sup>2</sup>Our analyses of the DA-related patent citation graph are standard, and can be similarly applied to the paper citation graph; cf. [27] [8]. And, just to be clear, while we examine research outputs such as papers (or patents), our study is undertaken precisely because “papers  $\neq$  impact”.

approximately \$7.56B total revenue in the last four quarters reported at [63], and is dominated by a few major companies (currently Synopsys, Cadence and Mentor Graphics, whose combined market capitalization today is approximately \$17.3B). Hundreds of smaller companies fight over a remaining single-digit percentage of the total EDA market. Figure 2 shows the evolution over time of industry demographics compiled by the market research firm Gary Smith EDA (GSEDA) [4]. Figure 3(a) shows that very few of the hundreds of EDA companies offer more than one product, and that even the “middle tier” of companies with between two and five products has been shrinking. Figure 3(b) shows the total numbers of tools offered in each of the main industry sectors according to the annual GSEDA “Wall Charts”.<sup>3</sup> Figure 4 shows EDA and semiconductor industry revenues over the past two decades: EDA has been stable at just over 2% of semiconductor billings.<sup>4</sup>

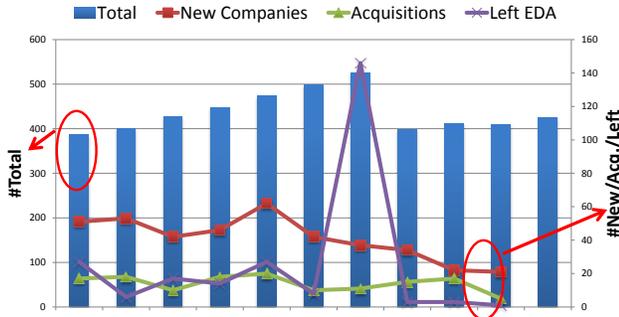


Fig. 2: Historical views of the EDA industry: total number of EDA companies, and new / acquired / closed companies, per year [4].

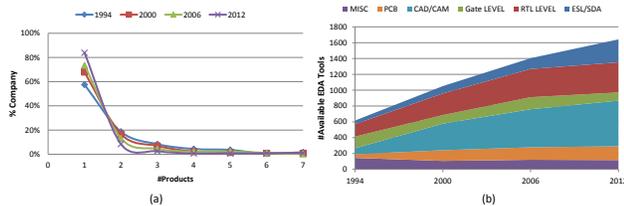


Fig. 3: (a) Percentage of EDA companies, plotted against the number of distinct products (occurrences in “Wall Charts”) per company. (b) Number of tools available in each “Wall Chart” segment from 1994 to 2012. Source: [4].

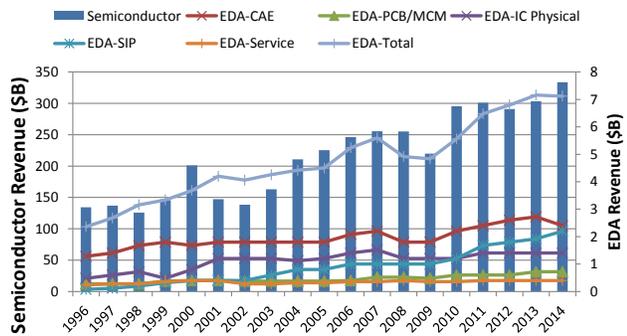


Fig. 4: Annual semiconductor industry revenues (blue bars) [64], and EDA industry total and per-segment revenues [63].

### Semiconductor industry-provided support for DA research.

The U.S.-based Semiconductor Research Corporation [59] has

<sup>3</sup>In the GSEDA Wall Charts, the categories of CAD/CAM, CAE-RT-level, CAE-Gate-level and Other have remained consistent over the years. PCB and system-level (SDA, ESL) have not always been separated out; for these categories, we obtain the reported segmentation from supporting data provided by [4].

<sup>4</sup>We cite EDA industry summary data compiled by both GSEDA and EDAC’s Market Statistics Service (MSS). There are differences, e.g., GSEDA reports \$6.3B industry revenue for 2014, versus \$7.56B reported by MSS [63]. (MSS includes the Semiconductor Intellectual Property (SIP) segment, which is not included by GSEDA.) Our reconciliation of the MSS and GSEDA breakdowns is available at [65].

since 1982 supported academic research in both design- and manufacturing-related science areas. According to [20], the SRC Global Research Collaboration (GRC) fraction of core funding going to the design areas (Computer-Aided Design and Test Sciences (CADTS) and Integrated Circuit and Systems Sciences (ICSS), at roughly equal levels) has increased over the past ten years from about 40% to 45%; see Figure 5. During 2004-2014, the CADTS split across Verification, Test and LPD (Logic and Physical Design tools) was initially steady at 25%-25%-50% but is now closer to 30%-30%-40%, while the ICSS funding split has been roughly even between system design and circuit design. In both areas there has been an increase in analog project content.<sup>5</sup> We consider below whether project portfolios of, e.g., NSF or SRC can be correlated to DA research impacts.

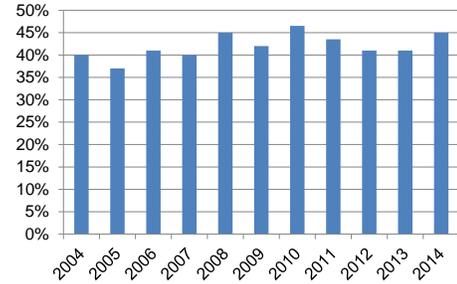


Fig. 5: Fraction of SRC’s core funding allocated to design areas [20].

### III. METHODOLOGY

Figure 6 summarizes our flow of data collection, processing and analyses – leading to a basis for DA research impact metrics. We now describe details of (i) research data collected and (ii) analyses performed.

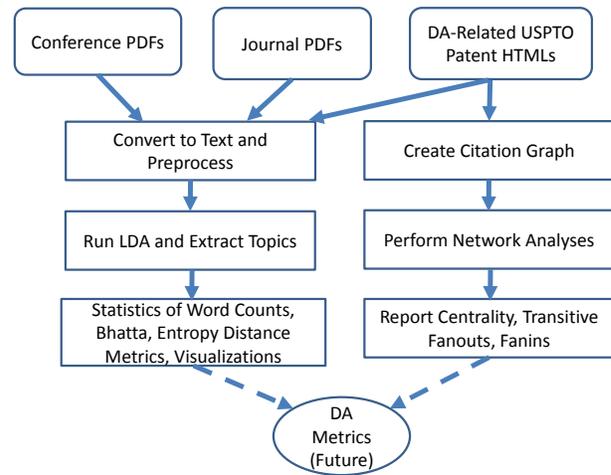


Fig. 6: Overall data collection and analysis flows.

#### A. DA Research Data Collected

**DA-related papers.** We analyze a collection of papers from DA conferences and journals, chiefly those sponsored by the IEEE

<sup>5</sup>Additionally, there have been four joint SRC-NSF programs in the design areas: MSET (Mixed Signal Electronic Technologies, 2001-2004), MCDA (Multi-core Design and Architecture, 2009-2012), FRS (Failure-Resistant Systems, 2013-2016) and STARSS (Secure, Trustworthy, Assured and Resilient Semiconductors and Systems, 2014-).

CEDA and ACM SIGDA professional societies.<sup>6</sup> Figure 7 shows timelines of the conferences and journals, and of the number of DA papers and patents.<sup>7</sup> The conference and journal papers themselves are near-universally available in searchable PDF form; exceptions are processed with OCR software, and then all papers are converted to text using *pdftotext*. We end up with text files for 47000+ conference and journal papers.<sup>8</sup>

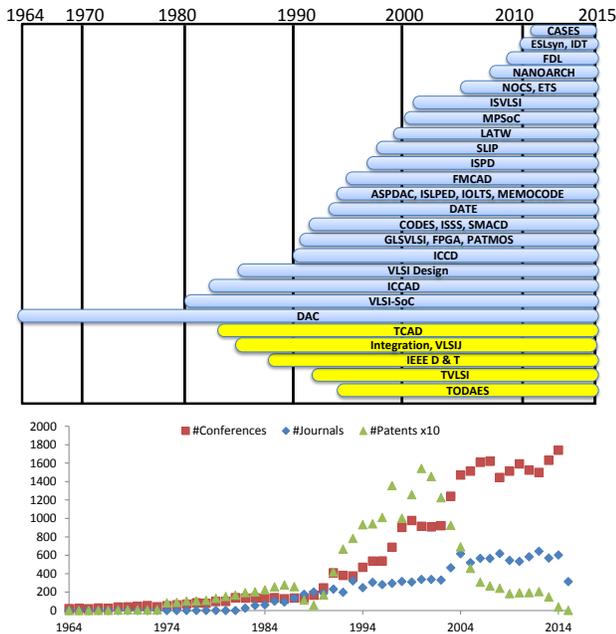


Fig. 7: (above) Timeline of EDA conferences and journals that we analyze, based on sponsorship lists supplied by IEEE CEDA [42] and ACM SIGDA [10]. (below) Per-year breakdown of the corpus of research outputs that we analyze.

**DA-related patents.** We obtain DA-related U.S. patents as full-text HTML from the USPTO website ([www.uspto.gov](http://www.uspto.gov)). We create the patent corpus as follows. We start from a “core” (or, level 0), and perform three levels of “forward” search in the directed graph of patent citations. We further perform one level of “backward” search from patents in the union of levels 0 and 1.<sup>9</sup> In the resulting patent

<sup>6</sup>Our collection is compiled from personal collections of ACM SIGDA’s annual compendia, DA Library, 25th Anniversary DVD; the DAC 40th Anniversary DVD; individual copies of conference proceedings; and online sources. IEEE CEDA-sponsored conferences and workshops include [42] ASP-DAC, DAC, DATE, ESLsyn, ESWEEK, ETS, FDL, FMCAD, GLSVLSI, ICCAD, IDT, IOLTS, ISVLSI, LATW, MEMOCODE, NoCS, PATMOS, SMACD, VLSI Design, and VLSI-SoC. ACM SIGDA-sponsored conferences and workshops include [10] ASP-DAC, DAC, DATE, ESWEEK, FPGA, GLSVLSI, ICCAD, ISLPED, ISPD, MEMOCODE, NANOARCH, NoCS and SLIP. The primary relevant society journal publications are IEEE TCAD, IEEE TVLSI, IEEE D&T, and ACM TODAES. We omit conferences and workshops without published proceedings (IWLS, TAU, MPSoC) and for the present study also omit circuits-centric meetings such as ISCAS or SBCCI. We include the historically significant Elsevier publication *Integration, the VLSI Journal*. Number of papers per source per year in our collection is given at [65]. Last, due to human error and the deadline for this paper, our reported analyses include only LATW 2015 papers and FPGA 1994–2003. Corrected and future updated studies will be maintained at [65].

<sup>7</sup>The growth in the number of papers is much faster than that of the EDA industry itself, perhaps reinforcing the “papers  $\neq$  impact” caveat above. The sharp decrease in the number of patents may be due to study methodology, economic climate, or increasing cost of patent pursuit combined with decreased value of software patents.

<sup>8</sup>We recognize that many influential works are not “papers” or “patents”. E.g., SPICE/HSPICE (e.g., [37]) are mentioned in 7200+ papers (over 15% of all papers!), the *International Technology Roadmap for Semiconductors* [17] is mentioned in 2700+ papers, the Weste-Eshraghian textbook [52] is mentioned in 4500+ papers, SIS and Sentovich [44] are mentioned in 1800+ papers, etc. Influential papers not from the specific conferences and journals analyzed are also missed by our analysis. Our ongoing work includes extensions to more complete views of DA research and knowledge.

<sup>9</sup>If patent A cites patent B, we say that there is a “forward” directed edge from vertex B to vertex A (i.e., B is a parent of A with respect to ‘knowledge’) in the corresponding directed graph of citations. One level of “backward” search from a given vertex (patent) obtains all patents that are cited by the given patent.

citation graph, a *source* node corresponds to a patent that does not cite any other patent, and a *sink* node corresponds to a patent that is not cited by any other patent. (We illustrate the terminologies of “forward” and “backward” searches, source and sink nodes, and levels in the citation graph in Figure 14(a) below.) Our corpus is the set of all patents found by this process. The full-text HTML file returned by the USPTO website for a given patent X provides all patents cited by patent X (edges incoming to (backward from) X), and all patents that cite patent X (edges outgoing from (forward from) X). At each of the three levels of the forward search, we traverse all edges outgoing from unexpanded vertices in the current collection; this is breadth-first search when the citation graph is treated as a directed graph. Our starting “core” consists of 4000 DA-related patents: (i) all patents with assignee  $\in$  {Cadence Design Systems, Mentor Graphics, Synopsys, Atrenta, Magma Design, Monterey Design, Springsoft, Gateway Design Automation, Valid Logic}, and (ii) a set of several hundred early U.S. patents in the semiconductor design and design automation field [21].<sup>10</sup> Our final patent corpus has 759507 patents. A list of these patent numbers is given at [65]. Removal of all HTML tags yields a text-based corpus. **NSF project abstracts and SRC research needs.** The U.S. National Science Foundation (NSF) has for decades supported DA research. We use NSF’s advanced award search function to find all projects under “CISE/CCF” with either of two well-known program managers for DA research, Dr. Sankar Basu and Dr. Robert B. Grafton. The query “CCF” AND “Sankar Basu” returns 729 funded projects from 1996 to 2015. The query “CCF” AND “Robert B. Grafton” returns 461 projects from 1984 to 2006. We export search results to a text file of all project abstracts, and this file is further broken down by year to enable more precise “alignment” studies of the project abstract corpus against the research literature. The SRC maintains Research Needs documents which are typically posted along with calls for proposal white papers in various programs. We obtained from SRC [55] a set of 36 Research Needs documents spanning 2000–2013, in the long-standing areas of circuit design, integrated system design, logic and physical design, verification and test.<sup>11</sup>

## B. Data Processing and Analysis Methods

**Text preprocessing.** After text extraction from PDFs of conference and journal papers, and from HTMLs of patents, we perform preprocessing to remove strings that have no meaning (e.g., those due to OCR artifacts for older papers; these typically have just two or three characters), punctuation marks and other anomalous strings. We filter out words with fewer than five letters unless the words are in a “whitelist”. Last, we filter out words that are in a “blacklist”. The whitelist includes DA-specific words and abbreviations that have four or fewer letters (e.g., cell, wire, ATPG, BIST, BDD, etc.). The blacklist contains common English words (e.g., articles, conjunctions, prepositions, etc.) as well as non-differentiating words that are ubiquitous in DA papers and patents

<sup>10</sup>These early intellectual properties are due to such entities as IBM, Honeywell, the U.S. Army and U.S. Air Force, Texas Instruments, Bell Telephone Labs, etc. – and their contents are often fascinating to read. E.g., in what is today known as the “SP&R” space, example titles (USPTO numbers), all with filing dates no later than the 1980s, include: “Process for Selecting Circuits with Optimum Power and Area Requirements” (T935003), “Automated Logic Mapping System” (T940008), “Automated System and Method for Partitioning and Mapping Circuit Units Onto Modules Including an Iterative Process” (T944001), “Method of Minimizing the Interconnection Cost of Linked Objects” (3617714), “Machine Process for Assigning Interconnected Components to Locations in a Planar Matrix” (3629843), “Element Placement System” (3654615), “Machine Process for Positioning Interconnected Components to Minimize Interconnecting Line Length” (3681782), “Method of Characterizing Critical Timing Paths and Analyzing Timing Related Failure Modes in Very Large Scale Integrated Circuits” (4698587), “Method for Optimizing Signal Timing Delays and Power Consumption in LSI Circuits” (4698760), “Static Timing Analysis of Semiconductor Digital Circuits” (4924430), etc.

<sup>11</sup>The 36 documents that we analyze also include several unique instances, e.g., Long Horizon Topics (2012), mixed-signal (2001), and circuit and system design (2002).

(e.g., circuit, technology, algorithm, average, automation, design, etc.). The whitelist and blacklist that we use are posted at [65].

**LDA analysis of papers and patents.** We perform latent Dirichlet analysis (LDA) on preprocessed text files. Relevant terminology, after [7], is as follows.

- A *dictionary*  $D^*$  is a “universe” of  $V$  distinct words  $w_1, \dots, w_V$ .
- A *document* is a sequence of words, each an element of  $D^*$ , which we treat as a multiset (“bag of words”).
- A *corpus* is a collection of documents.

Given a text corpus  $\mathbf{D}$ , we perform LDA analysis to obtain  $K$  topic vectors. A topic vector is a unit-length  $V$ -dimensional vector  $\mathbf{z} = \{z_1, \dots, z_V\}$ , in which  $z_v$  is proportional to the appearance frequency of the word  $w_v \in D^*$ . In examples that we show below (e.g., Figure 8), for each topic vector we show only the words corresponding to the  $J$  largest values of  $z_v$ . That is, we typically illustrate the results of LDA topic modeling as a set of  $K$  topics (e.g.,  $K = 12$ ), and  $J$  words for each topic (e.g.,  $J = 10$ ).

We use the publicly-available LDA implementation from *lda* [58]. Figure 8 shows the first ten words (i.e., with highest weight in the given vector’s distribution) in each of the top-12 topic vectors returned by LDA for the “four sisters” EDA conferences (DAC, ICCAD, DATE and ASPDAC), in the time intervals 1964-1970, 1980-1985, 1995-2000, and 2010-2015.<sup>12</sup> Truncation of words is due to stemming performed by the LDA implementation. “Hand labels” have been added for clarity.

**Quantification of distances between LDA topic models.** We compare the  $K$  topic vectors from two corpora to understand similarities and evolutions of topic models across different years or sources. Given two topic vectors  $\mathbf{z} = \{z_1, \dots, z_V\}$  and  $\mathbf{z}' = \{z'_1, \dots, z'_V\}$ , we define the *entropy distance* [53] between them as

$$l^{ENTROPY}(\mathbf{z}, \mathbf{z}') = \sum_{v=1}^V z_v \cdot \ln\left(\frac{z_v}{z'_v}\right) + \sum_{v=1}^V z'_v \cdot \ln\left(\frac{z'_v}{z_v}\right) \quad (1)$$

We also define the *Bhattacharyya distance* (Bhatta) [5] as

$$l^{BHATTYA}(\mathbf{z}, \mathbf{z}') = -\ln\left(\sum_{v=1}^V \sqrt{z_v \cdot z'_v}\right) \quad (2)$$

Given two corpora of words  $D$  and  $D'$ , we perform LDA analysis to obtain  $K$  topic vectors ( $K$  a user-specified parameter of the analysis) for each corpus. That is, we obtain topic vectors  $\mathbf{z}_k$  and  $\mathbf{z}'_{k'}$ , for  $1 \leq k, k' \leq K$ . We then calculate the distance  $l^\gamma(\mathbf{z}_k, \mathbf{z}'_{k'})$  between each pair of topic vectors ( $\mathbf{z}_k, \mathbf{z}'_{k'}$ ) from the two corpora (i.e., for all  $1 \leq k, k' \leq K$ ), where  $\gamma \in \{BHATTYA, ENTROPY\}$ . We then perform minimum-cost perfect matching in the bipartite graph of pairs ( $\mathbf{z}_k, \mathbf{z}'_{k'}$ ), with edge costs being the calculated distances. That is, we minimize

$$\begin{aligned} L^\gamma(\mathbf{D}, \mathbf{D}') &= \sum_{k, k'=1}^K \alpha_{k, k'} \cdot l^\gamma(\mathbf{z}_k, \mathbf{z}'_{k'}) \\ \text{subject to } &\sum_{k'=1}^K \alpha_{k, k'} = 1, \quad \forall k = 1, \dots, K \\ &\sum_{k=1}^K \alpha_{k, k'} = 1, \quad \forall k' = 1, \dots, K \end{aligned} \quad (3)$$

where  $\alpha_{k, k'}$  is a 0-1 indicator of matching between  $\mathbf{z}_k$  and  $\mathbf{z}'_{k'}$ . We take  $L^\gamma(\mathbf{D}, \mathbf{D}')$  as the distance between the corpora  $\mathbf{D}$  and  $\mathbf{D}'$ .

Given the above definitions, Figure 9 shows the pairwise entropy [53] and Bhatta [5] distances for sets of 12 LDA topic model vectors, computed for corpora of “four sisters” conference papers in four separate time intervals. With each metric, a smaller distance value indicates higher similarity. Because the two distance

<sup>12</sup>The term “four sisters” refers to the long-standing “special relationship” among these four conferences which, beyond their leading status in the DA field, share society sponsors and explicitly align their calendars to avoid conflicts.

metrics behave similarly, in the following we use and report only Bhatta distance. The figure also shows the Bhatta distance between the 12-vector LDA topic model for the “four sisters” conferences of 2013-2015, and the corresponding topic models for each preceding year. The relatively monotone behavior of this distance measure is intuitively reasonable.<sup>13</sup>

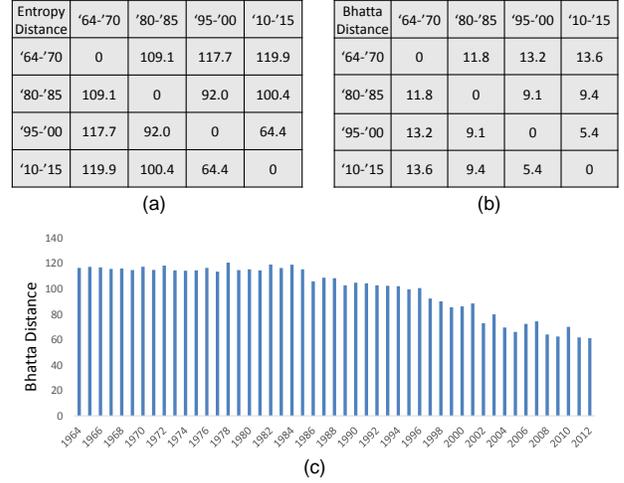


Fig. 9: (a) Entropy distance [53] and (b) Bhatta distance [5] between sets of 12 LDA topic vectors computed for four time intervals of the “four sisters” conferences. (c) Bhatta distance between the set of 12 LDA topic vectors for the “four sisters” conferences 2013-2015, and the corresponding sets of 12 LDA topic vectors for each preceding individual year.

**Betweenness centrality for citation graph analysis.** To analyze the impact of patents on EDA and its related fields (e.g., computer hardware, architecture, etc.), we use the widely used betweenness centrality measure [30]. Given a connected graph  $G = (V, E)$ , we compute the betweenness centrality of each vertex  $v \in V$  as follows.

$$C_b(v) = \sum_{s \neq v \neq t} \frac{p_{s,t}(v)}{p_{s,t}}, \quad \forall s, t \in V, s \neq t \quad (4)$$

where  $s, t \in V$  are the source and target vertices in the graph that are different from  $v$ ,  $p_{s,t}(v)$  is the number of shortest paths from  $s$  to  $t$  that pass through  $v$  and  $p_{s,t}$  is the number of shortest paths from  $s$  to  $t$ . For the citation graph, betweenness centrality measures the extent to which a vertex impacts and connects related fields.

## IV. ANALYSES AND RESULTS

### A. LDA Topic Models

1) *Stasis vs. Change in LDA Topic Models:* Figure 10 shows, for the ICCAD conference paper corpus, the Bhatta distance between a given year (“Year”) and two (resp. three) years earlier (“Year - 2” (resp. “Year - 3”). In most cases, the “Year - 3” distance is larger than the “Year - 2” distance, as would be expected. In years where the “Year - 2” and “Year - 3” distances are both relatively large and relatively different (2014, 2009, 2001), we might infer that the subject matter of ICCAD conference papers was rapidly evolving over the several years (2011-2014, 2006-2009, 1998-2001). On the other, in years where the two distance measures are nearly the same and relatively small (1987, 1993, 2007, 2008), we might infer that the ICCAD subject matter exhibited “stasis” over the several years (1984-1997, 1990-1993, 2004-2007, 2005-2008).<sup>14</sup>

<sup>13</sup>Years that show non-monotonicity, such as 1996 and 2001, may indicate noteworthy dynamics in the DA research world. See Section IV.A.1 below.

<sup>14</sup>We further observe that in the years 1988, 1990, 2009, 2011 and 2014, the “Year - 2” distance is larger than the “Year - 3” distance. This could indicate “transients” of research topics or technical emphases in the conference that appeared for two years but then fell into the background. Of course, such speculations demand more careful, detailed study.

| 1964 - 1970   | 1980 - 1985   |
|---|---|
| Topic 1: [framework] control graphic display machin process perform direct repres block complet<br>Topic 2: [programming] languag statement compil symbol variabl regist devic array control assign<br>Topic 3: [mechanics] point symbol coordin surfac draft plotter geometr prepar layout light<br>Topic 4: [circuit simulation] network matrix voltg branch nonlinear integr linear transistor differenti transient<br>Topic 5: [analysis] engin manufactur analysi detail product process effect produc chang increas<br>Topic 6: [graph theory] graph branch order plane connect planar determin common subgraph group<br>Topic 7: [PWB] wire compon connect board placement interconnect layout print assign point<br>Topic 8: [enterprise] peopl industri thing question answer train studi million great labor<br>Topic 9: [test] gate fault failur reliabl detect diagnost propag common delay network<br>Topic 10: [floorplan] pin modul packag partit entri signal block class compat primari<br>Topic 11: [analog] variabl paramet model respons analysi determin current calcul accelr frequenc<br>Topic 12: [numerical] mechan connect order problem requir space discuss relat final optim | Topic 1: [routing] channel wire vertic connect track layer segment horizon grid router<br>Topic 2: [circuit simulation] process devic model paramet capacic extract analysi perform region transistor<br>Topic 3: [high-level] languag relat databas repres model implement describ represent specif syntheis<br>Topic 4: [layout] layout block symbol array gate connect automat process power physic<br>Topic 5: [numerical] model techniqu analysi element integr digit evalua waveform current dynam<br>Topic 6: [graph theory] graph optim minim order reduc partit matrix variabl implement repres<br>Topic 7: [UI] softwar graphic engin process display interact technol support requir applic<br>Topic 8: [physical design] modul placement compon wire connect locat global partit process assign<br>Topic 9: [test] fault detect pattern vector testabl observ techniqu combin faulti control<br>Topic 10: [architecture] control hardwar regist memori execut modul processor instruct process signal<br>Topic 11: [device] gate delay signal model transistor network clock cmos propag connect<br>Topic 12: [algorithm] control function comput digital number connect ratiou equir algorithm tructur |
| 1995 - 2000   | 2010 - 2015   |
| Topic 1: [power] power estim activ switch transit consumpt energi dissip voltg gate<br>Topic 2: [analysis] model signal error analysi perform paramet measur estim process analog<br>Topic 3: [circuit simulation] matrix model linear order approxim frequenc induct respons vector moment<br>Topic 4: [placement] partit graph placement cluster optim connect block assign vertic minim<br>Topic 5: [architecture] process model softwar hardwar interfac communic implement applic embed architectur<br>Topic 6: [system] schedul memori execut perform processor optim instruct applic architectur resourc<br>Topic 7: [high-level] control regist signal syntheis behavior model modul implement clock instruct<br>Topic 8: [routing] wire segment layer channel grid crossstalk track steiner router grid<br>Topic 9: [device] transistor current layout devic voltg cmos interconnect coupl model effect<br>Topic 10: [timing] delay gate optim clock wire retim buffer signal critic insert<br>Topic 11: [verification] variabl boolean repres transit order bdd decomposit symbol cover implement<br>Topic 12: [test] fault detect vector sequenc pattern sequenti testabl propos bist procedur | Topic 1: [circuit simulation] model estim propos sampl perform error analysi paramet matrix random<br>Topic 2: [security] control sensor secur signal attack measur monitor implement digit requir<br>Topic 3: [memory] memori access block instruct write perform propos regist address architectur<br>Topic 4: [physical design] placement layout pattern wire graph optim propos pin assign connect<br>Topic 5: [embedded] model hardwar implement process applic softwar architectur perform compon accelr<br>Topic 6: [architecture] schedul execut applic perform resourc thread processor parallel model process<br>Topic 7: [test] error fault detect reliabl failur techniqu defect probabi propos correct<br>Topic 8: [power] power energi voltg current consumpt switch leakag suppli propos reduc<br>Topic 9: [high-level] variabl model propret syntheis implement order boolean formal repres transit<br>Topic 10: [NOC] network noc router packet communic channel latenc buffer interconnect architectur<br>Topic 11: [timing] delay clock gate signal optim propos buffer critic stage perform<br>Topic 12: [thermal] temperatur thermal model devic tsv transistor technol variat stress effect   |

Fig. 8: 12-topic LDA models (top 10 words shown) for years 1964-1970, 1980-1985, 1995-2000 and 2010-2015 of the “four sisters” conferences. “Hand labels” are given in square brackets.

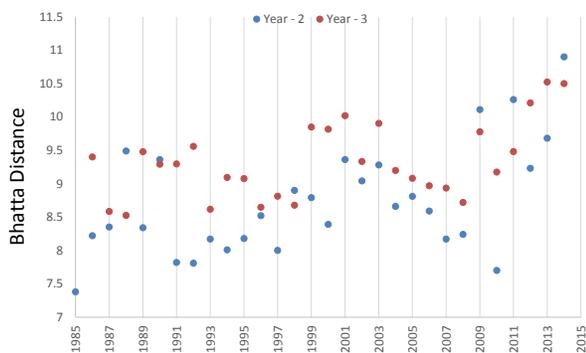


Fig. 10: Bhatta distance between ICCAD papers of a given year and ICCAD papers of two or three years earlier.

2) *Alignment (Lead-Lag) Analyses:* Figure 11 (red bars) shows alignment analyses of leading or lagging relationships between NSF project abstract and “four sisters” conference text corpora. Parts (a) and (b) of the figure respectively show normalized Bhatta distances of conference papers in three-year windows, from the 1994-1996 and 2011-2012 NSF project abstracts. Parts (c) and (d) of the figure repeat the analysis using two-year windows, with distances respectively calculated from the 1995-1996 and 2010-2011 windows. The blue bars in parts (d) and (b) of the figure repeat these analyses using SRC needs document corpora from 2011-2012 and 2009-2013, respectively. As explained above (recall Figure 9), smaller distance values imply greater similarity. While absolute variations in the distance metric are small, we observe that distances decrease in the years after the projects are initiated, as one would expect if the NSF projects impact the trajectory of the field.<sup>15</sup>

3) *Visualizations by Word Clouds and Incidence Curves:* We perform simple word count-based visualization of the DA research corpora using *word clouds* [60]. A word cloud shows the most frequently occurring words, with font size positively correlated with frequency of occurrence. Our word cloud plotting is done using the R language, with package *wordcloud* [61]. Figure 12 shows word clouds of the text corpora from the “four sisters” conferences in 1980-1985 and 2010-2015. As expected, this visualization matches

<sup>15</sup>Figure 11(a) might reveal an interesting possibility, namely, that (i) a field might move into a given state, (ii) this state is then reflected in a “zeitgeist” (e.g., NSF funding of projects), and (iii) the state is then reaffirmed by the research project outcomes. This brings up various “chicken vs. egg” or “credit and impact assessment” challenges, as we note in our conclusions below. A further comment is that the particular result of Figure 11(a) may reflect potential “stasis” of the research literature in the early and mid-1990s, which we have noted in the context of Figure 10 above.

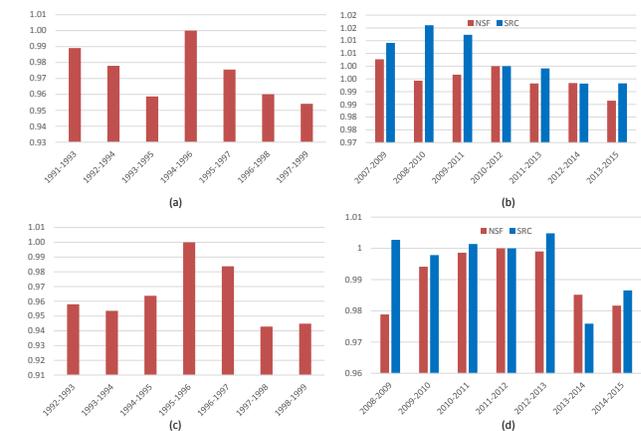


Fig. 11: Red bars (a)-(b): Alignment of top-12 topics from NSF project abstracts in the three-year intervals 1994-1996 and 2010-2012, respectively, with displacements of up to three years in the “four sisters” conference paper corpus. Bhatta distance is used, with smaller distance suggesting stronger correlation or alignment. (c)-(d): Similar alignment calculation using two-year intervals 1995-1996 and 2011-2012, respectively. Blue bars repeat (d) and (b) analyses using 2011-2012 and 2009-2013 SRC needs documents.

well with Figure 8.

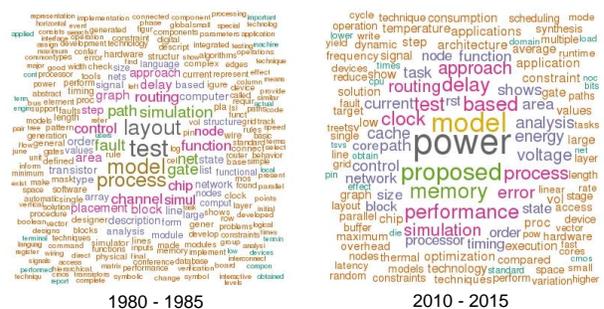


Fig. 12: Word clouds of subsets of the text corpus for the “four sisters” conferences: (a) 1980-1985 and (b) 2010-2015.

The number of research works in which a given term appears can also indicate emerging, (e.g., “new topic”), declining (e.g., “solved problem”), “dead end”), or stable levels of attention to a particular challenge or field. Figure 13(a) shows *incidence curves* (cf. “hype cycles”, a term popularized by Gartner [62]) of eight DA-related terms in our corpus of conference and journal papers. We ignore case and search for the number of papers per year that contain these terms. For the recent physical design challenge of “multiple-patterning”, we count papers that use the union of terms “{double, multi, triple} patterning”. Figure 13(b) illustrates the potential as of

incidence curves to assess latencies, in this case between conference and journal corpora. In the figure, incidence curves for most terms (e.g., “BDD”) typically peak two years later in journals than in conferences, with the exception of “DVFS”, for which both peaks occur in 2013.<sup>16</sup>

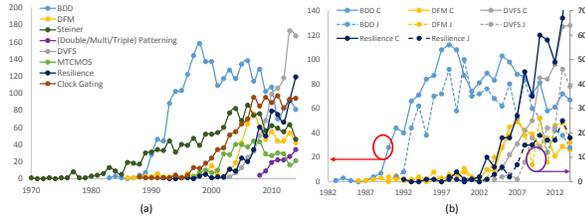


Fig. 13: Incidence curves (a) of selected terms, in all 47000+ papers; (b) of several of these terms, in conference (solid) and journal (dotted) papers (note the two y-axis scales).

### B. USPTO Patent Citation Graph Analyses

As described above, we create a corpus of 759507 USPTO patents. The citation graph over these patents contains 4717209 edges (citations between patents in the corpus), 99290 source (does not cite any other patent) vertices, and 490079 sink (is not cited by any other patent) vertices.<sup>17</sup> We plot size vs. frequency of fanout cone size in Figure 14(b) and observe a long-tailed distribution: a few vertices have large transitive fanouts (possibly interpretable as high impact in the field), whereas a large number of vertices have small transitive fanouts.<sup>18</sup> Figure 14(c) shows a similar long-tailed distribution for betweenness centrality. This suggests that only a few vertices (patents) may be linking adjacent fields. For example, the two patents with highest betweenness centrality are US6076734 (“Methods and systems for providing human/computer interfaces”, cone size = 3060) and US6558320 (“Handheld personal data assistant (PDA) with a medical device and method of using the same”, cone size = 2258); these respectively seem to link design automation and human-computer interface, and design automation and biomedical instrumentation). US6558320 appears in the citation graph because it cites patent US5885245, which is a descendant of the DA-related patent US4813013.

### V. TOWARD MEASUREMENTS OF RESEARCH IMPACT

Our discussion has noted several ways to view and analyze DA research activity. However, these are still only potential building blocks toward *metrics* of DA research *impact*. We remain far from a magic formula that scores research impact fairly and that can be used, e.g., for early identification of high-impact or high-value research directions. Moreover, many basic challenges (e.g., separating “correlation” from “causation”) are untouched by the approaches described above.

In this section, we note three examples of quantifiable impacts that should be at least partly attributable to DA research activity: (i) design productivity improvement, (ii) company valuation, and (iii) commercial tool availability. We suggest that these examples

<sup>16</sup>Three comments. (i) A caveat for DVFS is that this is still a “relatively recent” term. (ii) The DVFS example may point to decreasing latencies between conference and journal publication, and/or increasing “substitutability” of conference and journal publications. (iii) For the example of “resilience”, the incidence curve seems to point to an as-yet unseen peak of journal paper activity.

<sup>17</sup>We include isolated vertices in the sink count.

<sup>18</sup>The maximum (resp. average) sizes of the transitive fanout cones of source vertices are 8412 (resp. 5497). The two patents with largest fanout cones in our graph are US4451895 (“Interactive computer aided design system”, cone size = 8412) and US4492956 (“Graphics display system and method including preclipping circuit”, cone size = 8406). The two patents with largest outdegree are US5892900 (“Systems and method for secure transaction management and electronics rights protection”, outdegree = 2057) and US5077607 (“Cable television transaction terminal”, outdegree = 1504). US5077607 appears in the citation graph because it cites US4451895, and a number of patents on television and consumer electronics cite US5077607.

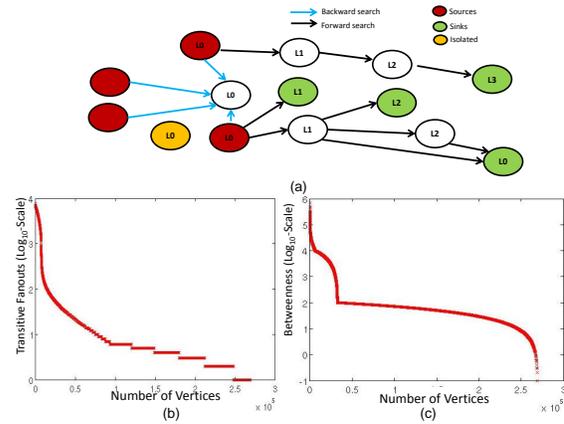


Fig. 14: (a) An illustration of a citation graph, showing core nodes (level 0), forward and backward searches, and source and sink nodes. Distributions of (b) fanout cone size and (c) betweenness centrality in the citation graph over 759000+ DA-related U.S. patents. We omit sinks and isolated vertices from the analyses of transitive fanout and betweenness centrality.

can provide early calibrations and sanity-checking for any proposed methodologies for measuring DA research impact.

### A. Impact: Design Productivity Improvement

Ever since SEMATECH’s 1993 “design productivity gap” chart (cf. Figure 2 and accompanying discussion in [22]), the impact of DA innovation on design productivity has been a focus for the semiconductor industry. Design productivity has been a “grand challenge” throughout the entire existence of the ITRS roadmap, and the ITRS Design Cost Model, initiated in 2001 [24] [17], proposes quantified impacts of specific DA advances on design productivity. In this subsection, we show how a case study of IBM POWER microprocessors, based on published papers, can quantify design productivity improvements over multiple product generations.

The IBM POWER processor family has a 20+ year history since the 1990 RISC-based POWER1. Table I shows the evolution of the IBM POWER processor. For each product, Year = year of introduction, Freq = operating frequency (GHz), #Cores = number of cores, #Tx = number of transistors, Die size = die area in  $mm^2$ , and Highlights = unique technical highlights of the given product.<sup>19</sup> Operating clock frequency peaked in the POWER6 generation, after which POWER7 and POWER8 reduced frequency while emphasizing scale-out systems and applications. The wide range of supply voltage levels across operating modes, which is seen in the latter generations, could not have been achieved without advances in physical design, timing analysis and power management methodologies. Integration of eight quad-threaded cores in POWER7 brought a huge leap in the number of transistors. Figure 15 shows nearly 80% growth in number of *synthesized* macros between POWER6 and POWER8. Interestingly, the total number of macros decreased significantly during this time, indicating that the size of synthesizable macros increased tremendously. Assuming that the design team size stays almost constant (on the order of hundreds [28]), these data suggest that

<sup>19</sup>Some details of the Highlights are as follows. (i) POWER6 [14] more than doubled the operating frequency of POWER5 [11] at constant power; to meet the power challenge, it deployed a dynamic power management system that enabled operation at clock frequencies > 5GHz in high-performance applications, as well as operation at < 100W in power-sensitive applications. (ii) POWER7 [51] integrated 8 cores per die, targeting scale-out systems. The use of integrated eDRAM for the L3 cache achieved greater density (3× that of 6T SRAM) with less power consumption, as predicted in the System Drivers chapter of the 2001 ITRS (MPU driver model). (iii) POWER8 [13] was the first to employ resonant clocking in a main clock mesh system. Dubbed a microprocessor “for big data”, it integrated on-chip accelerators for cryptography, memory compression, and a coherent interface (CAPI) for external hardware acceleration. Thus, external IO bandwidth increased significantly.

POWER processor designs have leveraged synthesis more heavily in recent generations. Again, without the consistent advancements in EDA technology, these design productivity improvements would not be possible.

TABLE I: IBM POWER Processor Designs

|            | P4 [3]          | P5 [11] | P6 [14]               | P7 [51]          | P8 [13]          |
|------------|-----------------|---------|-----------------------|------------------|------------------|
| Year       | 2001            | 2004    | 2007                  | 2010             | 2014             |
| Freq (GHz) | 1.0-1.3         | 1.65    | 4 - 5                 | 3.5 - 4          | 3 - 4+           |
| #Cores     | 2               | 2       | 2                     | 8                | 12               |
| #Tx        | 174M            | 276M    | 790M                  | 1.2B             | 4.2B             |
| Die size   | 412             | 389     | 341                   | 567              | 650              |
| Node (SOI) | 180nm           | 130nm   | 65nm                  | 45nm             | 22nm             |
| Highlights | First Dual-Core | SMT     | High Freq, Power Mgmt | Multicore, eDRAM | Res. Clock, CAPI |

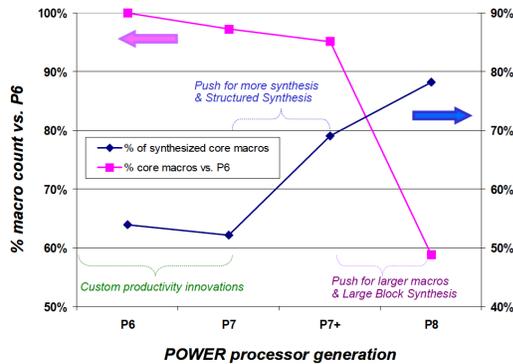


Fig. 15: Synthesized macro counts in IBM POWER processors [57].

Returning to the MPU model and roadmap in the ITRS System Drivers chapter, we see that the 1.4x per node frequency scaling predicted in 2001 was oblivious to platform power limits, which were added into the roadmap in 2007. Faced with power, manufacturability and yield challenges, multicore architecture enabled by continued density scaling have become the focal point for performance scaling. Design technology advances such as power management and resonant clocking are another, “equivalent scaling” thrust for semiconductor products. Emerging devices predicted by ITRS, such as eDRAM, enable a further dimension of design-based scaling as well.

The ITRS Design Chapter has since 2001 maintained a Design Cost model for leading-edge MPU and SOC products. The previous paper [24] explicitly calls out the challenge of design productivity and design cost explosion back to early 2000s. However, significant improvement on design productivity successfully defends for the prosperity of semiconductor and allows the benefit of technology scaling to be fully utilized. Design complexity and design technology milestones for the IBM POWER processor family are shown in Figure 15. We observe that Large Block Reuse (which, according to [24], first reached production in 2000) dominates the design hierarchy to reduce design effort.

### B. Impact: M&A Valuation

Over the past several decades, the three largest EDA companies (Synopsys, Cadence and Mentor) have been party to many merger and acquisition transactions. In the EDAC “IC Physical Design & Verification” category (\$1.5B/year total revenues), which includes synthesis, place and route, the history of acquisitions (with year and estimated post-earnout total value, if known) includes the following examples [26]:

- Cadence acquisitions = Tangent (1988), HLD Systems (1996, \$303M), Cooper & Chyan (1996, \$836M), Ambit (1998, \$510M), CadMOS (2001, \$201M), Silicon Perspective (2001, \$598M), Plato (2002, \$85M), Simplex (2002, \$676M), Celestry (2003, \$146M), Get2Chip (2003, \$160M), Verplex (2003, \$214M);

- Mentor acquisitions = Sierra DA (2007), Ponte (2008), Pyxis Technology (2012), OASYS (2013);
- Synopsys acquisitions = Gambit DA (1998), Everest DA (1999, \$82M), Avant! (2000, \$1925M), Monterey (2004), TeraRoute (2007), Magma DA (2011, \$507M), Extreme DA (2011).

Company valuations in such transactions very likely embody some impacts of DA research: if not the recoding of papers and algorithms, then (i) the graduate training of capable engineers, and possibly (ii) reuse of technology starting points from public research artifacts (Capo, PRIMA, FastCap, minisat, CUDD, HotSpot, SPICE, ...). More explicit transfers of DA research outcomes (e.g., via startups or IP licenses) have quantifiable valuations as well. On the other hand, the history of SP&R startups and acquisitions given above also points out how strongly *individuals* affect overall impacts. How to separate the impacts of DA research from those of people and business factors will be challenging.

### C. Impact: Commercial EDA Tools

Similarly, commercial EDA tools address a perceived available market, and typically must offer some kind of differentiated technology. Thus, commercial EDA tools can also be a form of “DA research impact”. However, our preliminary studies contradict somewhat the folklore wisdom of a “seven to 10 years” lag between research papers and deployment in production tool flows. Speculatively, this might be because research needs and technology roadmaps are both generated by the semiconductor companies, and reach academics and the EDA industry at the same time.

We briefly consider the recent examples of power- and temperature-driven EDA tooling, as well as cloud-based EDA. By 2001, the ITRS [17] Design Chapter had already highlighted power and reliability as overarching “grand challenges” for design technology. Further, the ITRS Design Cost model (2001-) and Low-Power Design Technology model (2009-) have set out specific prescriptions for design productivity (e.g., symmetric and asymmetric multiprocessing) and low-power design techniques [17] [46]. Such influences motivate DA research activity (both directly and indirectly through funding consortia), as well as EDA startup and R&D efforts. A partial picture is given by Figure 16(a), which shows incidence curves for low power- and temperature-related keywords,<sup>20</sup> and Figure 16(b), which shows the emergence of commercial power, thermal, and cloud EDA tools. We do not see the expected lag between research papers and commercial tool deployment for such “grand challenge” topics as low-power design, temperature- and reliability-aware design, and rearchitecting of algorithms for the cloud. This again suggests that for some real-world impacts, DA research as it is traditionally measured might not be a proximate cause.

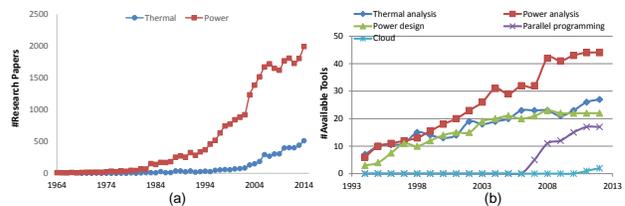


Fig. 16: (a) Incidence curves of low power- and temperature-related terms. (b) Emergence of power, thermal and cloud EDA tools, from GSEDA Wall Charts [4].

### D. Conclusions and Further Analyses

As noted at the outset, metrics of DA research impact could help identify high-value research directions and results, formulate higher-impact research programs with limited resources, or establish the ROI of DA research (hopefully, to justify increased investment).

<sup>20</sup>Numbers plotted reflect the union of papers containing “low power” or “low-power”, and the union of papers containing “thermal-aware”, “temperature-aware”, “thermal-constrained” or “temperature-constrained”.

We have discussed several measurements and analyses (LDA-based topic models, topic model alignment and temporal latency analyses, citation graph analyses) of DA research *outputs* (papers, patents, EDA companies) upon which future metrics of DA research *impact* might be based. Our studies suggest that the research literature can experience periods of stasis and of rapid evolution; that lead/lag relationships among various types of research outcomes (e.g., papers and commercial tools) have chicken-egg or other complex dynamics; etc. While we do not achieve “metrics of impact”, we do recognize that papers and patents by themselves are not “impact”; we further suggest that design productivity, M&A valuations, and commercial tool deployment are among potential “ultimate measures” of DA research impact within the IC design space. We plan to pursue follow-on work such as (i) analyses of additional citation graphs (papers, papers+patents, etc.), (ii) retrospective assessment of early, assumed indicators of impact (e.g., “best paper” vs. “test of time”), (iii) development of statistical and machine-learning models for real-world impacts of both individuals and individual research results, and (iv) development of predictors of future high-impact DA research.

#### ACKNOWLEDGMENTS

We thank Shishpal Rawat and the IEEE CEDA board for initiating and supporting this project, and ACM SIGDA for additional help. Laurie Balch of Gary Smith EDA (GSEDA) generously provided historical EDA market share and “wall chart” data. Kevin Lepine, Kathy Emblar and Sandy Owens of MP Associates generously compiled session and exhibit data from DAC and ICCAD conferences in support of several early analyses. Puneet Sharma and Xu Xu [23] wrote original USPTO scraping scripts, and Vasudev Vikram wrote many additional data collection and analysis scripts that we used. We thank SRC’s Bill Joyner and David Yeh, and IBM’s John Darringer and David Kung, for valuable inputs. Last, many students at UCSD, UT Austin and U. Virginia contributed to the data compilation. In particular, Wei-Ting Jonas Chan, Kwangsoo Han, Jiajia Li and Hyein Lee contributed valuable data cleaning, analysis and figure generation efforts to this paper. ABK would like to dedicate this paper to the memory of Mr. Gary Smith (March 9, 1941 - July 3, 2015), whose support of – and impact on – the EDA field cannot be measured.

#### REFERENCES

- [1] R. Alghamdi and K. Alfalqi, “A Survey of Topic Modeling in Text Mining”, *I. J. ACSA* 6(1) (2015), pp. 147-153.
- [2] Y. An, J. Janssen and E. E. Milios, “Characterizing and Mining the Citation Graph of the Computer Science Literature”, *KIS* 6(6) (2004), pp. 664-678.
- [3] C. J. Anderson, et al., “Physical Design of a Fourth-Generation POWER GHZ Microprocessor”, *Proc. ISSCC*, 2001, pp. 232-233.
- [4] L. Balch, Gary Smith EDA (<http://www.garysmitheda.com/>), *personal communication*, August 2015.
- [5] A. Bhattacharyya, “On a Measure of Divergence between Two Multinomial Populations”, *Sankhya* 7(4) (1946), pp. 401-406.
- [6] D. M. Blei and J. Lafferty, “Correlated Topic Models”, *Proc. NIPS*, 2006, pp. 147-154.
- [7] D. M. Blei, A. Y. Ng and M. I. Jordan, “Latent Dirichlet Allocation”, *J. MLR* 3 (2003), pp. 993-1022.
- [8] L. Bolelli, S. Ertekin and C. L. Giles, “Clustering Scientific Literature Using Sparse Citation Graph Analysis”, *Proc. KDD*, 2006, pp. 30-41.
- [9] L. Cao and F.-F. Li, “Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes”, *Proc. ICCV*, 2007, pp. 1-8.
- [10] A. Cozzi, *personal communication*, August 2015.
- [11] J. Clabes, et al., “Design and Implementation of the POWER5 Microprocessor”, *Proc. DAC*, 2004, pp. 670-672.
- [12] T.-K. Fan and C.-H. Chang, “Exploring Evolutionary Technical Trends from Academic Research Papers”, *Proc. IWDAS*, 2008, pp. 574-581.
- [13] E. Fluhr, et al., “POWER8: A 12-Core Server-Class Processor in 22nm SOI with 7.6Tb/s Off-Chip Bandwidth”, *Proc. ISSCC*, 2014, pp. 96-97.
- [14] J. Friedrich, et al., “Design of the Power6 Microprocessor”, *Proc. ISSCC*, 2007, pp. 96-97.
- [15] T. L. Griffiths and M. Steyvers, “Finding Scientific Topics”, *PNAS*, 101(1) (2004), pp. 5228-5235.
- [16] T. Hofmann and J. Puzicha, “Learning from Dyadic Data”, *Proc. NIPS*, 1999, pp. 466-472.
- [17] *International Technology Roadmap for Semiconductors*, <http://www.itrs.net> and <http://www.itrs2.net>, particularly the Design and System Drivers chapters, 1998-present.
- [18] S. Jameel and W. Lam, “An Unsupervised Topic Segmentation Model Incorporating Word Order”, *Proc. SIGIR*, 2013, pp. 203-212.
- [19] Y. Jo, C. Lagoze and C. L. Giles, “Detecting Research Topics via the Correlation between Graphs and Texts”, *Proc. KDD*, 2007, pp. 370-379.
- [20] W. Joyner and D. Yeh, *personal communication*, August 2015.
- [21] A. B. Kahng, *personal archives*.
- [22] A. B. Kahng, “Design Technology Productivity in the DSM Era”, *Proc. ASP-DAC*, 2001, pp. 443-448.
- [23] A. B. Kahng, P. Sharma and X. Xu, “Structural Analysis of the USPTO Patent Citation Graph”, *unpublished manuscript*, 2004.
- [24] A. B. Kahng and G. Smith, “A New Design Cost Model for the 2001 ITRS”, *Proc. ISQED*, 2002, pp. 190-193.
- [25] T. Kakkonen, N. Myller and E. Sutinen, “Applying Latent Dirichlet Allocation to Automatic Essay Grading”, *Natural Language Processing* 4139 (2006), pp. 110-120.
- [26] T. Katsioulas, *personal communication*, August 2015.
- [27] “A Parser for Google Scholar”, <http://www.icir.org/christian/scholar.html>
- [28] D. Lammers, et al., “IBM flexes Power4 as it ships Regatta servers”, *EETimes News & Analysis article*, 2001, [http://www.eetimes.com/document.asp?doc\\_id=1144041](http://www.eetimes.com/document.asp?doc_id=1144041).
- [29] A. Levi, O. Mokryn, C. Diot and N. Taft, “Finding a Needle in a Haystack of Reviews: Cold Start Context-based Tool Recommender System”, *Proc. Recom. Sys.*, 2012, pp. 115-122.
- [30] L. Leydesdorff, ““Betweenness Centrality” as an Indicator of the “Interdisciplinarity” of Scientific Journals”, *J. of ASIST* 58(9) (2007), pp. 1303-1319.
- [31] F.-F. Li, “Bag of Words Models”, *CVPR Short Course*, 2007, [http://vision.cs.princeton.edu/documents/CVPR2007\\_tutorial\\_bag\\_of\\_words.ppt](http://vision.cs.princeton.edu/documents/CVPR2007_tutorial_bag_of_words.ppt).
- [32] D. J. MacKay and L. C. B. Peto, “A Hierarchical Dirichlet Language Model”, *J. NLE* 1(3) (1995), pp. 289-308.
- [33] J. McAuley and J. Leskovec, “Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text”, *Proc. Recom. Sys.*, 2013, pp.165-172.
- [34] J. D. McAuliffe and D. M. Blei, “Supervised Topic Models”, *Proc. NIPS*, 2008, pp. 121-128.
- [35] T. Minka, “Estimating a Dirichlet Distribution”, *Technical Report MIT*, 2000.
- [36] R. M. Nallapati, A. Ahmed, E. P. Xing and W. W. Cohen, “Joint Latent Topic Models for Text and Citations”, *Proc. KDD*, 2008, pp. 542-550.
- [37] L. W. Nagel and D. O. Pederson, “SPICE: Simulation Program with Integrated Circuit Emphasis”, *Memorandum ERL-M382*, Univ. of California, Berkeley, 1973.
- [38] K. Nigam, A. K. McCallum, S. Thrun and T. Mitchell, “Text Classification from Labeled and Unlabeled Documents Using EM”, *Machine Learning* 39(2-3) (2000), pp. 103-134.
- [39] A. M. Popescu, B. Nguyen and O. Etzioni, “OPINE: Extracting Product Features and Opinions from Reviews”, *Proc. HLT*, 2005, pp. 32-33.
- [40] L. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth and M. Welling, “Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation”, *Proc. KDD*, 2008, pp. 569-577.
- [41] H. Qi, B. Chen, J. Pei, B. Qiu, P. Mitra and C. L. Giles, “Detecting Topic Evolution in Scientific Literature: How Can Citations Help?”, *Proc. LKM*, 2009, pp. 957-966.
- [42] S. Rawat, *personal communication*, August 2015.
- [43] A. Sangiovanni-Vincentelli, “The Tides of EDA” *IEEE D&T*, 20(6) (2003), pp. 59-75.
- [44] E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton and A. Sangiovanni-Vincentelli, “SIS: A System for Sequential Circuit Synthesis”, *Technical Report UCB/ERL M92/41*, Univ. of California, Berkeley, 1992.
- [45] Z. Y. Shen, J. Sun, Y.-D. Shen, “Collective Latent Dirichlet Allocation”, *Proc. ICDD*, 2008, pp. 1019-1024.
- [46] G. Smith, “Updates of the ITRS Design Cost and Power Models”, *Proc. ICCD*, 2014, pp. 161-165.
- [47] M. J. Wainwright and M. I. Jordan, “Graphical Models, Exponential Families, and Variational Inference”, *J. FTML* (11-2) (2008), pp. 1-305.
- [48] H. M. Wallach, “Topic Modeling: Beyond Bag-of-Words”, *Proc. ICML*, 2006, pp. 977-984.
- [49] X. Wang and E. Grimson, “Spatial Latent Dirichlet Allocation”, *Proc. NIPS*, 2007, pp. 1577-1584.
- [50] X. Wang and A. McCallum, “Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends”, *Proc. KDD*, 2006, pp. 424-433.
- [51] D. Wendel, et al., “The Implementation of POWER7: A Highly Parallel and Scalable Multi-Core High-End Server Processor”, *Proc. ISSCC*, 2010, pp. 102-103.
- [52] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective*, Addison-Wesley, 1985.
- [53] A. T. Wilson and D. G. Robinson, “Tracking Topic Birth and Death in LDA”, *technical report SAND2011-6927*, Sandia National Laboratories, 2011.
- [54] S. Xu, L. Zhu, X. Qiao, Q. Shi and J. Gui, “Topic Linkages between Papers and Patents”, *Proc. ICAST*, 2012, pp. 176-183.
- [55] D. Yeh, *personal communication*, September 2014.
- [56] L. Zhang, Y. Guo, X. Chen, W. Shao and L. Chen, “Classification of Topic Evolutions in Scientific Conferences”. <http://sei.pku.edu.cn/~yaoguo/papers/Zhang-SEKE-14.pdf>
- [57] M. Ziegler, et al., “POWER8 Design Methodology Innovations for Improving Productivity and Reducing Power”, *Proc. CICC*, 2014, pp. 1-9.
- [58] “lda: Topic modeling with latent Dirichlet Allocation”, <http://pythonhosted.org/lda/>
- [59] Semiconductor Research Corporation, <http://www.src.org/>
- [60] “Tag cloud” entry in Wikipedia, [https://en.wikipedia.org/wiki/Tag\\_cloud](https://en.wikipedia.org/wiki/Tag_cloud)
- [61] “wordcloud” R package online (author: I. Fellows), <https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf>
- [62] Gartner “Hype Cycles”, <http://www.gartner.com/technology/research/hype-cycles/>
- [63] EDAC Market Statistics Service, [http://edac.org/initiatives/mss/newsletter\\_q1\\_2015](http://edac.org/initiatives/mss/newsletter_q1_2015) (see also: [http://www.edac.org/sites/default/files/users/mss/MSS\\_2015\\_Category\\_Definitions\\_FINAL.pdf](http://www.edac.org/sites/default/files/users/mss/MSS_2015_Category_Definitions_FINAL.pdf))
- [64] SIA “Semiconductor Industry Billing History”, [http://www.semiconductors.org/industry\\_statistics/historical\\_billing\\_reports/](http://www.semiconductors.org/industry_statistics/historical_billing_reports/)
- [65] DA Metrics website, <http://vlsicad.ucsd.edu/DA-METRICS/>