

# Curriculum Vitae

## **Kian-Lee TAN, Professor**

Department of Computer Science  
National University of Singapore  
Computing 1, 13 Computing Drive  
Singapore 117417

URL: <http://www.comp.nus.edu.sg/~tankl>

Email: tankl@comp.nus.edu.sg

## **Fellow, Singapore-MIT-Alliance (2000/01-2005/06)**

## **Senior Faculty Member, NUS Graduate School for Integrative Science and Engineering (2002-2010)**

## **Education**

Kian-Lee TAN received his B.S.(First-class Hons), M.S. and Ph.D. in computer science, from the National University of Singapore, in 1989, 1991 and 1994 respectively.

## **Research Interests**

Kian-Lee's major research interests include query processing and optimization (centralized, multiprocessor and distributed systems), indexing and organization of data in advanced database systems (spatial databases, multimedia databases, sequence databases) and information retrieval, database performance, wireless data dissemination, database security, and peer-to-peer computing. He also has interests in computational biology, in particular in protein 3D structure analysis and comparison and gene expression data mining and clustering.

## **Experiences**

Kian-Lee was a Visiting Scientist at IBM's Almaden Research Center, San Jose, California, from January~92 to July~92. He worked with the High Performance Relational Database Management Systems group in introducing partitioning properties into the Starburst Extensible Database Management System. He also participated in discussions on parallelizing Starburst.

He was also invited to spend a year (from June~94 to June~95) with the Distributed Database Group at CSIRO's Division of Information Technology, Australia. There, he was involved in the design and prototyping of the query processor for the Virtual Database System, an instance of a Virtual Information System. He also participated in the design and prototyping of the SQL-like frontend for the Spatial Database Management Tool SiroDBMS developed at CSIRO.

Kian-Lee was a Senior Scientist at the Genome Institute of Singapore from 1 July 2001 to 30 June 2003.

## Awards

Kian-Lee received the 1997/1998 Outstanding University Researchers Award. He has also been nominated for the award in 2003 and 2005.

## Publications

Kian-Lee has published over 230 papers in international conferences and journals. Some of these have appeared in prestigious conferences (e.g., SIGMOD, VLDB, ICDE) and journals (e.g., ACM Transactions on Databases, The Very Large Data Bases Journal, IEEE Transactions on Knowledge and Data Engineering). He has also co-authored a tutorial on Parallel Query Processing (by IEEE CS Press), a monograph on Advanced Indexing Techniques (by Kluwer Academic Publisher) and a monograph on Wireless Data Dissemination (by Kluwer Academic Publisher). See <http://www.comp.nus.edu.sg/~tankl/publications.html>.

## Company

Co-founded Geofoto Pte Ltd (in 1999), and is currently a non-executive director of Geofoto (since 2001). Geofoto Pte Ltd is a spin-off from NUS (as a result of our research in the VIPER project on content-based multimedia retrieval) that provides content management solutions. Geofoto Pte Ltd has recently (in 2004) been renamed to Thothe Technologies Pte Ltd (<http://www.thothe.com>).

## Patents

- “Apparatus and method for performing transformation-based indexing of high-dimensional data”, Beng Chin Ooi, Kian-Lee Tan, Stephane Bressan, Cui Yu, US Patent US 7,117,217 B2 (issued on 3 October 2006); Singapore Patent P-No 103262 (issued on 30 July 2004).
- “Transformation-based method for indexing high-dimensional data for nearest neighbour queries”, Cui Yu, Beng Chin Ooi, Kian-Lee Tan, US Patent P-No 6,834,278 issued on 21 December 2004.

## Code/Executable Releases

Here are some of the source code/executables that we have released to the community (algorithms developed are also available upon request):

- [iDistance](#): An efficient and elegant high-dimensional index for kNN search
- [GiMP](#): A framework for quick implementation of known transformation-based indexes
- [NSteg](#) - A Steganographic File System on Linux kernel 2.4.
- [WmXML](#) - A watermarking system for XML Data. A GUI system created in Visual C++.
- BestPeer/PeerDB: Planning to commercialize the system. Code/executables are available for academic use upon request.
- BATON\* (running on PlanetLab): A tree based structured network that achieves

- the search at  $O(\log_M N)$  -- to be released soon.
- [Genesis](#): A suite of algorithms on 3D protein structure alignment, classification and protein
- BLAST++: This is a tool that is integrated with NCBI BLAST, allowing multiple, say  $K$ , queries to be searched against a database concurrently. Code/executables are available for academic use upon request.

### **Administrative Duties**

#### **National**

- Member, R&D Working Group of the Cirrus Cloud Computing Testbed, (Sept 08 – present)
- Member, Public Sector Funding (PSF) Panel of the Science and Engineering Research Council (SERC) (1 March 2009 – 28 Feb 2010)

#### **University**

- Chair, Task Force on Computational Biology (To study the need for an undergraduate program) (Feb 2003-Aug 2003)
- Member, Task Force on Bioinformatics Master Program (To study the need for a Master program). (Nov 2003)
- URC Panel (2006-2008)

#### **School of Computing**

- Assistant Dean, School of Computing (Sept 2000 – June 2001).
- Deputy Head, Department of Computer Science (June 2001 – July 2009).
- Member, Faculty Promotion and Tenure Committee (1 July 2008 – 30 June 2010).
- Co-Chair, Task Force on Undergraduate Computational Biology Program (To design a curriculum for the program).
- Chair, SoC Undergraduate Computational Biology Program Advisors (July 2003-Jun 2006)
- Member, SoC Journal/Conference Ranking Committee (July 1999-present)

#### **Service to Industry**

- Regularly review project proposals for SIIRD (Singapore-Israel Industrial Research and Development Foundation) under EDB.

### **Professional Activities**

#### **Editorial Board**

- The VLDB Journal: The International Journal on Very Large Data Bases, Springer (2007-2013)
- IEEE Transactions on Data and Knowledge Engineering, IEEE CS Press (Feb 2009-Jan 2011)
- International Journal on Information Technology, World Scientific Publishers (2000-2005)

- International Journal on Web Engineering and Technology, Inderscience Enterprises Ltd (Since 2003).
- International Journal on Mobile Information Systems, IOS Press (since 2004).
- Journal of Computing Science and Engineering, KIISE (since 2007)
- Editorial Review Board Member of the Journal of Database Management, Idea Group Publishing (since 2001).
- Editorial Review Board Member of the PVLDB (Proceedings of the VLDB Endowment).
- Area Editor (Distributed Database Systems). Encyclopedia of Database Systems, Springer, 2008.

#### **Guest Co-Editors**

- Distributed and Parallel Databases (DAPD) on Peer Based Data Management (2009).
- IEEE TKDE on Peer Based Data Management 16(7) (2004).
- MONET on Mobile Data Management 8(4) (2003).

#### **Professional Societies**

- Member, Executive Committee, IFIP TC 5 Specialist Group on Bioinformatics (2004-2007)
- Member, Steering Committee, International Conference on Mobile Data Management (2002-2009)
- Secretary, Steering Committee, International Conference on Database Systems for Advanced Applications (2005-2010)

#### **Conference Program Committee Member**

- International Conference on Management of Data (SIGMOD'05, SIGMOD'07, SIGMOD'09 (Group Leader))
- International Conference on Very Large Data Bases (VLDB'98, VLDB'02, VLDB'04, VLDB'06, VLDB'08)
- International Conference on Data Engineering (ICDE'99, ICDE'03, ICDE'04, ICDE'05 Vice-Chair (OLAP, Data Warehouse, Statistical DBMS Track), ICDE'07 (TPC), ICDE'07 (Demo), ICDE'08, ICDE'09, ICDE'10 Vice Chair)
- International Conference on Extending Database Technology (EDBT'04)
- International Conference on Scientific and Statistical Database Management (SSDBM'09)
- International Symposium on Spatial and Temporal Databases (SSTD 2009)
- International Symposium on Database Systems for Advanced Applications (DASFAA'99, DASFAA'01, DASFAA'03, DASFAA'04, DASFAA'09)
- International World Wide Web Conference (WWW'07, WWW'08, WWW'09, WWW'09 Poster Track)
- International Symposium on Parallel and Distributed Computing (ISPDC'05)
- Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'03)

- International Conference on Visual Database Systems (VDB'2002)
- International Workshop on Multimedia Databases (MMDB'03, MMDB'04)
- International Conference on Electronic Commerce and Web Technologies (EC-Web'00, EC-Web'01, EC-Web'02, EC-Web'03, EC-Web'04, EC-Web'05, EC-Web'06, EC-Web'07)
- International Conference on Mobile Data Management (MDM'01, MDM'03, MDM'04, MDM'06, MDM'07, MDM'08, MDM'09)
- International Conference on Parallel and Distributed Systems (ICPADS'00, ICPADS'05)
- International Conference on Data Warehousing and Knowledge Discovery (DaWaK '05)
- International Conference on Service-oriented Computing (ICSOC'05, ICSOC'06, ICSOC'07)
- International Conference on Advanced Information Networking and Applications (AINA'07, AINA'08, AINA'09)
- International Conference on Parallel Processing (ICPP'03)
- International Workshop on Agents and Peer-to-Peer Computing (AP2PC'03, AP2PC'04, AP2PC'05, AP2PC'06, AP2PC'07, AP2PC'08)
- International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE'08)
- International Workshop on Ranking in Databases (DBRANK'09, DBRANK'10)
- International Workshop on Data Management for Sensor Networks (DMSN'06, DMSN'07)
- International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P'05, DBISP2P'06, DBISP2P'07)
- International Conference on Web Engineering (ICWE'05, ICWE'06)
- International Conference on Advances in Mobile Multimedia (MoMM'2005, MoMM'2006, MoMM'2007)
- International Conference on Mobile and Ubiquitous Systems: Networks and Services (MobiQuitous'06)
- International Workshop on Self-Managing Database System (SMDB'05)
- International Conference on Embedded Software and Systems (ICESS'05)
- International Conference on Networks (ICON'04)
- Brazilian Symposium on Databases (SBBD'2004, SBBD'2005, SBBD'07)
- International Symposium on Applied Computing (SAC'03, SAC'04, SAC'05, SAC'07, SAC'08, SAC'09)
- International Conference on Information Systems and Engineering (ISE'01, ISE'03)
- International Workshop on Biological Data Management (BIDM'03, BIDM'04, BIDM'05)
- IEEE International Workshop on Multimedia Databases and Data Management (MDDM'06, MDDM'07)
- IEEE International Workshop on Scalable Stream Processing Systems (SSPS'07, SSPS'08)

- International Conference on Web Information Systems Engineering (WISE'00, WISE'02, WISE'04, WISE'07)
- International Conference on Database and Expert Systems Applications – (DEXA'08)
- International Conference on Embedded Software and Systems (ICCESS 2005)
- IEEE Workshop on Workflow and Data Flow for Scientific Applications (2005)
- Australian Database Conference (ADC'02, ADC,03, ADC'05, ADC'06, ADC'07, ADC'10)
- International Conference on Management of Data (COMAD 2005)
- Asia Pacific Web Conference (APWeb'04, APWeb'05, APWeb'06, APWeb'07, APWeb'08)
- IEEE Intl. Conference on Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC'08)
- International Workshop on Data Semantics in Web Information Systems (DASWIS'01, DASWIS'02)
- International Workshop on Cooperative Internet Computing (CIC'02, CIC'04, CIC'06)
- International Workshop on Mobility in Databases and Distributed Systems (MDDS'01, MDDS'04, MDDS'05)
- International Workshop on XML Schema and Data Management (XSDM'05, XSDM'06)
- International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE'08)
- International Workshop on Privacy, Security, and Trust in KDD (PinKDD'07, PinKDD'08)
- International Workshop on Secure Semantic Web (SSW'08)
- International Workshop on P2P Data Management, Security and Trust (PDMST'05, PDMST'06)
- International Workshop on Internet Data Management (IDM'99)
- International Database Symposium on Mobile, XML and Post-relational Databases (IDS'00)
- International Symposium on Database, Web and Cooperative Systems (DWACOS'99),
- International Workshop on Mobile Data Access (MDA'98)
- International Workshop on Knowledge Discovery and Temporal Reasoning in Deductive and Object-Oriented Databases (TR-DOOD'95)

#### **Conference Organizing Committee Member**

- Program (Co-) Chair
  - International Conference on Very Large Data Bases (VLDB'10) – Core DB Track
  - International Symposium on Database Systems for Advanced Applications (DASFAA'06)
  - International Conference on Mobile Data Management (MDM'02)

- Human.Society@Internet Conference (Singapore PC Chair) (HSI'01, HIS'03, HIS'05)
- Industrial Chair - International Conference on Conceptual Modelling (ER'98)
- Tutorial Co-chair - International Symposium on Database Systems for Advanced Applications (DASFAA'08)
- Demo Co-chair - International Symposium on Database Systems for Advanced Applications (DASFAA'10)
- Publication Chair - International Conference on Mobile Data Management (MDM'01)
- Finance Co-chair - Asia Pacific Software Engineering Conference (APSEC'00)
- Organizing Committee – Asia Pacific Bioinformatics Conference (APBC'05)

**Panel:**

- Panelists in DASFAA 2006 on “DB/IR Problems: What Wine, What Bottle”

**Reviewer**

- ACM Transactions on Database Systems (TODS)
- ACM Transactions on Sensor Networks (TOSN)
- ACM Transactions on the Web (TWeb)
- The Very Large Data Base Journal (VLDBJ)
- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- IEEE Transactions on Networking (TON)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- Distributed and Parallel Databases
- Data and Knowledge Engineering (DKE)
- Multimedia Tools and Applications (MTAP)
- ACM Multimedia (MM)
- ACM Computing Surveys
- Information Systems
- The Computer Journal
- Information Sciences
- Journal of Bioinformatics and Computational Biology (JBCB)
- Bioinformatics
- BMC Bioinformatics
- Knowledge and Information Systems
- Information and Software Technology
- International Journal on World Wide Web
- Electronics and Telecommunications Research Institute (ETRI) Journal
- ACM International Conference on Management of Data (SIGMOD)
- ACM Symposium on Principles of Databases (PODS)
- International Conference on Very Large Data Bases (VLDB)
- IEEE International Conference on Data Engineering (ICDE)
- International Conference on Information and Knowledge Management (CIKM)
- International Conference on Cooperative Information Systems (COOPIS)

- International Conference on Conceptual Modeling (ER)
- International Symposium on Advances in Spatial Databases (SSD)
- ACM International Conference on Multimedia (ACM-MM)
- International Conference on Multimedia Modeling (MMM)
- ACM International Conference on Geographic Information Systems (ACM-GIS)
- International Conference on Foundation on Data Organization (FODO)

## **Grants**

### *Funded by University (Principle Investigator)*

- R-252-000-307-112 Securing Streaming Data: Privacy and Access Control (Sep 07 – Aug 09). Amount: \$49,500.
- R-252-000-237-112 Distributed System for Standing Subscriptions (Jan 06 – Dec 08). Amount: \$260,000.
- R-252-000-228-112 UNOdb: Towards UNcheatable Outsourced Databases (Aug 05 – Dec 07). Amount: \$123,000.
- R-252-000-183-112 Authentication and Dissemination of Data in Edge Computing (Mar 04 – Feb 06). Amount: 67, 300.
- R-252-000-158-112 Peer-based Data Management (June 03 – May 05). Amount: \$83, 000.
- 981674 High-dimensional data search (RS augmentation grant) (July 1998-June1999). Amount: \$6,000.
- 982694 Evaluation and experimentation of efficient data retrieval strategies (Nov 1998-Jul 2001). Amount: \$49,000.
- 982705 Intelligent integration of information (RS augmentation grant) (May 1999-Nov 1999). Amount: \$6,000.

### *Research projects funded by government agencies or industry*

- Co-Director of Project C3: Fudan-NUS Competency Center for Peer-to-Peer Computing (Project No: 22.222.1006.158.08.0) supported by IDA (Dec 2002-Dec 2004). Amount: \$650,000.
- Principal Investigator: R-252-000-015/303 Strategic Programme on Computer Security funded by NSTB (Jan 1997 – Jun 2003 (Took over as PI from Oct 2000)). Amount: \$4,200,000.
- A collaborator in the research project on Digital Image and Video Album (DIVA) funded by NSTB (April 1999- June 2003). Amount: \$1,200,000

## **Teaching History: Summary of courses Taught**

- SMA5506: Database Technology (2001/02 -- 2005/06). This is a 40-hour lecture course was offered to (Master/PhD) students in the Singapore-MIT-Alliance (SMA) program.
- CS2102: Introduction to Database Systems (1996/97 -- 2000/01). This is the first course in databases for SoC students. It is a core module and has over 500 students per year.
- CS3223: Database Management Systems (1997/98, 1999/2000, 2000/01,

2002/03). This is the second course in databases for SoC students. It is an elective course (i.e., optional for students).

- CS5225: Parallel & Distributed Database Systems (1996/97, 1998/99, 2003/04 -- 2007/2008). This is an advanced level database course targeting Master (Part-time) and PhD students.
- CS5322: Database Security (2006/2007 – 2007/2008). This course is designed for Master and PhD students.
- CS5208: Foundations of Database Systems (2008/2009). This course is designed for Master and PhD students to build them up on the foundations of database systems
- CS5226: Database Tuning (2008/2009). This course is originally designed to teach student about database internals and how a database can be tuned to optimize performance. I have revamped it to study self-tuning methods (largely from industrial products). The course also has a number of practical labs that expose students to real database tuning experiences.

Fellow, Singapore-MIT-Alliance CS program (July 2001-June 2006)

Senior Faculty Member of NUS Graduate School for Integrative Science and Engineering (2003-2010).

### **Research Students Supervised**

#### ***Current PhD students***

- Cao Yu (Optimizing Complex Queries with Multiple Relational Instances)
- Cao Jianneng (Anonymizing Data Streams)
- Cheng Wei Wei (Secure Data Publishing)
- Wu Ji (Optimizing Data Stream Processing: It's All About Time)
- Wu Wei (SMA program) (Mobile and Pervasive Computing)
- Xiang Shili (Query Processing in Sensor Network)

#### ***Former students/postdoctoral fellows who have completed***

Postdoctoral fellows (5)

- Bin Cui (Mar 2004-Mar 2006) (Currently: Associate Professor, Peking University)
- Gao Cong (Mar 2004- Sept 2004) (Currently: Researcher, Microsoft China)
- Xuejian Xiong (2003-2004) (Currently: Postdoc at the University of Toronto/Samuel Lunenfeld Research Institute)
- Yan Wang (1999-2003) (Currently: Senior Lecturer, Macquarie University)
- Shichao Zhang (1998-1999) (Currently: Research Fellow, University of Technology, Sydney)

PhD students (11)

- Li Yingguang (2003-2009) – Keyword Search in Peer-to-Peer Systems (First Job: DBS Bank)

- Ji Liping, (2002-2007) – Mining Localized Co-expressed Gene Patterns from Microarray Data (First Job: Postdoc at Dept of Plant Biology, Stanford University, with Sue Rhee)
- Zhou Yongluan, (2002-2007) (President's Graduate Fellowship 2005) – Scalable and Adaptive Distributed Stream Processing (First Job: Researcher in the Distributed Information Systems Laboratory (LSIR) at EPFL)
- Zeyar Aung, (2001-2006) – Computational Analysis of 3D Protein Structures (First Job: Research Fellow at Institute for Infocomm Research)
- Shu Yanfeng, (2001-2006) – Supporting Complex Queries in P2P Networks (Currently: Research Scientist at CSIRO Tasmanian ICT Centre, Australia)
- Eng Pin Kwang, (2001-2005) – Skyline/Preference Query Processing (First Job: Research Fellow at Institute for Infocomm Research)
- Zhou Xuan, (2002-2005) (Dean's Award 2003 & 2004; President's Graduate Fellowship 2003 & 2004) – Steganographic File System (First Job: Research Fellow and Project Manager at L3S Research Center, Germany)
- Goh Shen Tat, (2001-2005) – Managing Cache for Efficient Query Processing (First Job: Research Fellow at Institute for Infocomm Research)
- Shou Lidan, (1998 – 2003) (Dean's Award 2001) (co-supervised with Dr Zhiyong Huang) - Querying Large Virtual Model for Walkthrough Applications (First Job: Lecturer at Zhejiang University, PRC)
- Manoranjan Dash, (1998-2001) – Efficient Clustering in High-Dimensional Databases (Currently: Assistant Professor, Nanyang Technological University)
- Cai Jun (1996-1999) – Wireless Information Retrieval (Currently working in industry in Singapore) (Currently: (Telecom) Industry in Singapore)

*MSc (by research) students (23)*

- Yu Feng (Towards an Internet Scale Stream Processing Service with Loosely-Coupled Entities)
- Ahmar Faraz (Prediction of Protein Structures) (with David Hsu)
- Ma Xi (Steganographic DBMS) (with HweeHwa Pang)
- Song Xuyang, 2005 (Selectivity Estimation in XML Query Processing) (with Chee Yong Chan)
- Zhang Han, 2004 (PeerStore: A Peer-to-Peer Backup System)
- Zhang Zonghong, 2004 (Mining Biclusters in Gene Expression Data)
- Chionh Chern Hooi, Jason, 2003 (Protein Structure: Alignment & Classification)
- Dong Shujie, 2003 (Indexing Moving Objects)
- Goh Chin Hong, 2003 (StegVault: A Steganographic P2P Data Bank)
- Qian Jinghui, 2003 (Steganographic B-Trees)
- Artem Vorobiev, 2003 (A Component Based Platform for P2P Applications)
- Zhang Ming, 2003 (Supporting Rich Queries in P2P Environment)
- Wu Xiu, 2003 (Collaborative Proxy System for Distributed Web Content Transcoding)
- Pang Xiaolin, 2002 (Secure Mobile Agent)
- Loh, Winston, 2002 (Secure Execution Environment for P2P Applications)

- Ruan Yixin, 2002 (Indexing Schemes for Walkthrough in Large Virtual Environments)
- Ong Twee Hee, Rick , 2002 (Similarity Search in Genomic Databases)
- Ren Jian, 2002 (Mobile Agent Based Internet Marketplaces)
- Cheong Chia Chou, 2000 (Processing and Optimization of Teams)
- Eng Pin Kwang, 2000 (Information Retrieval in Digital Marketplaces)
- Thiang Lay Foo, 1999 (Shape-based Image Retrieval)
- Goh Shen Tat, 1999 (A Multi-Feature Cluster-Based Image Retrieval System)
- Yee Chia Yeow, 1998 (Content-based Image Retrieval Techniques for Large Image Collections)

*Diploma (Exchange students doing the dissertation part of their diploma requirement with me) (3)*

- Radu Negoescu, 2001 (from Romania) (Securing XML Documents)
- Martin Landers, 2004 (from Munich) (Backup Store in P2P)
- Colin Rouse, 2005 (from University of Cape Town, South Africa) (Peer-based Data Management)