An empirical assessment on the adequacy of utilizing pre-grant publications in patent classification analysis

Max Kuan
Graduate Institute of Patent Research
National Taiwan University of Science and Technology

專利情報與資訊計量研究室
Patent Intelligence & Informetrics Lab.

Outline

- Background
  - Patent classification, patent classification analysis, pre-grant publication, motivation
- Methodology
  - Assumptions, data, analysis methods
- Result
- Conclusion

Background: Patent Classification

- Every patent is assigned **one or more classification symbols** during its application process by the patent authority based on one or more **classification schemes** such as
  - International Patent Classification (IPC)
  - Cooperative Patent Classification (CPC)
  - For US patents, US Patent Classification (USPC)
  - …

Background: Patent Classification

United States Patent **5,376,580**
Kish , et al. **December 27, 1994**

Wafer bonding of light emitting diode layers

Abstract

A method of forming a light emitting diode (LED) includes providing a temporary growth substrate that is selected for compatibility with fabricating LED layers having desired mechanical characteristics. For example, lattice matching is an important consideration. LED layers are then grown on the temporary growth substrate. High crystal quality is thereby achieved, whereafter the temporary growth substrate can be removed. A second substrate is bonded to the LED layers utilizing a wafer bonding technique. The second substrate is selected for optical properties, rather than mechanical properties. Preferably, the second substrate is optically transparent and electrically conductive and the wafer bonding technique is carried out to achieve a low resistance interface between the second substrate and the LED layers. Wafer bonding can also be carried out to provide passivation or light-reflection or to define current flow.

Inventors: **Kish; Fred A.** (San Jose, CA), **Steranka; Frank M.** (San Jose, CA), **DeFevere; Dennis C.** (Palo Alto, CA), **Robbins; Virginia M.** (Los Gatos, CA), **Uebbing; John** (Palo Alto, CA)
Assignee: **Hewlett-Packard Company** (Palo Alto, CA)
Family ID: **21889110**
Appl. No.: **08/036,532**
Filed: **March 19, 1993**

Current U.S. Class: **438/26**; 438/27; 438/28; 438/455
Current CPC Class: H01L 25/0756 (20130101); H01L 33/005 (20130101); H01L 33/0062 (20130101); H01L 33/0079 (20130101); H01L 33/0087 (20130101); H01L 33/30 (20130101); H01L 2924/0002 (20130101); H01L 33/145 (20130101); H01L 2924/0002 (20130101); H01L 2924/00 (20130101)
Current International Class: H01L 33/00 (20060101); H01L 021/20 ()
Field of Search: ;437/127,129,130,905,974,117,229 ;148/DIG.135

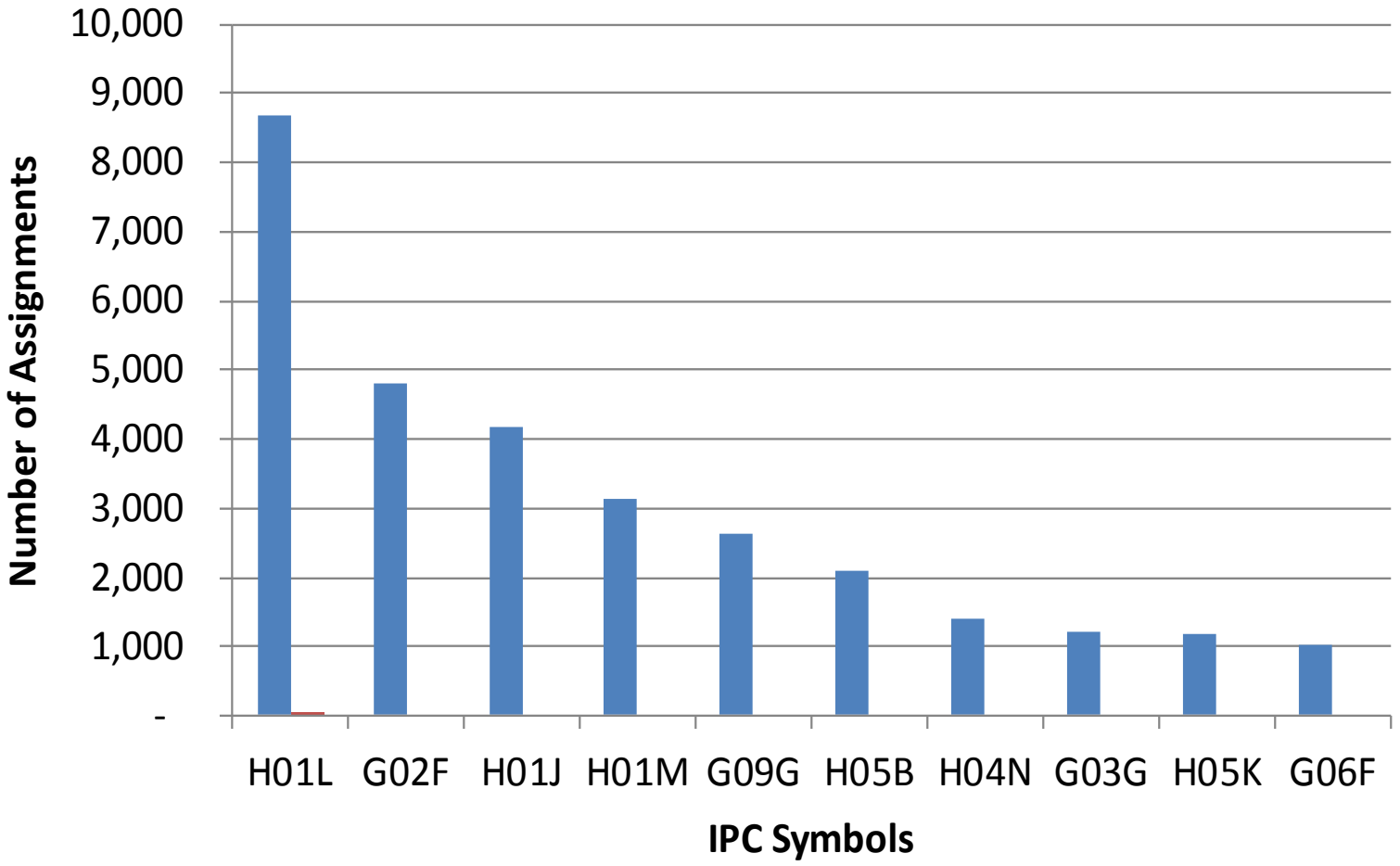
Background: Patent Classification

- All classification schemes provide a **tree-like hierarchical taxonomy of technology areas**
- For example, a USPC symbol **623/2.11**
  - class code **623** represents the technology area "prosthesis, parts thereof, or aids and accessories therefore
  - subclass code **2.1** represents a subordinate technology area "heart valve"
  - subclass code **2.11** represents a subordinate technology area "heat valves combined with surgical tool"

Background: Patent Classification Analysis (PCA)

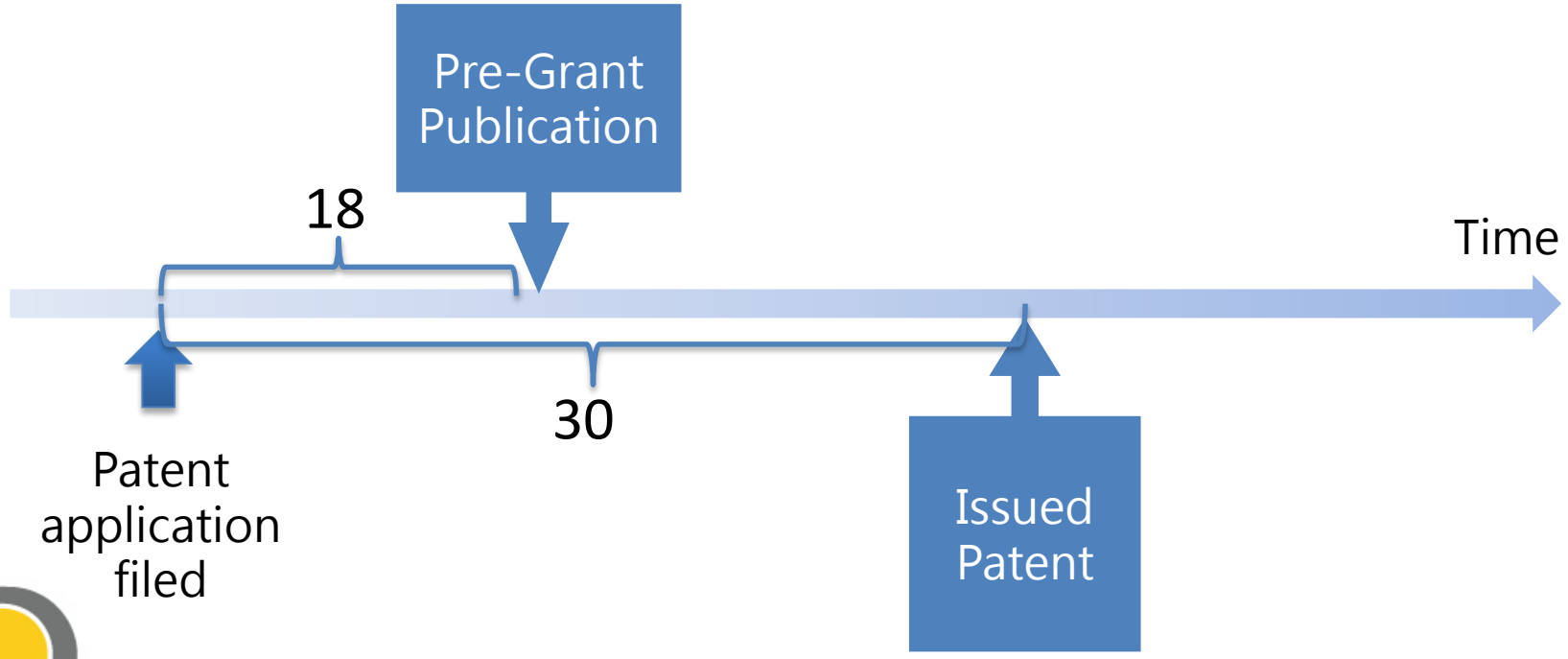
- A popular practice in patent analysis
  - All commercial patent analytic systems have built-in PCA function
- Usually used to investigate the **technology focuses** of an entity by looking at **what classification symbols are assigned most frequently** to its patent portfolio

Background: Patent Classification Analysis (PCA)



Background: Pre-grant Publication

- A pre-grant publication (PGPub) is a public document of an utility patent application published 18 months after filing



Background: Pre-grant Publication

- PGPub and its issued patent are **both** assigned classification symbols
  - The two are not necessarily identical
  - PGPub's classification is usually rough and is mainly for task routing
- Most PCAs use **classification symbols of issued patents**, meaning the analytic result actually reflects the reality 30 months earlier.

Motivation:
Can we use PGPub classification symbols to achieve trustworthy PCA while reducing the time lag down to 18 months?

Methodology: Assumptions & Data

- To make sure PGPub classification symbols are accurate enough
  - We assume that the classification symbols assigned to the corresponding issued patent are the “correct” ones.
  - We **compare the classification symbols of the PGPubs to those of the corresponding issued patents** to see how different they are.
- We collected about 235,000 **utility patents issued in the year 2012** from USPTO database and their PGPubs and compare their respective **USPC symbols**

Methodology: Analysis Methods

- Using a patent's classification symbols as example
  - **726/32**; 380/201; 705/57; 726/27; 726/31; 726/33
    - Bold-faced symbol (**main symbol**) is the representative one
- PCA of commercial systems usually collects classification symbols by three methods
  - Method **1**: Only the **class code of the main symbol: 726**
  - Method **2**: the **complete main symbol: 726/32**
  - Method **3**: the entire set of symbol: **726/32**; 380/201; 705/57; 726/27; 726/31; 726/33
- PCAs usually **do not consider the hierarchical relationship** among the symbols

Methodology: Analysis Methods

- Corresponding to the 3 PCA methods, we conduct three analyses
  - Analysis **1** compares the **main classification class codes**
  - Analysis **2** compares the **main symbols**
  - Analysis **3** compares **the sets of classification symbols**
- For all 3 analyses, we calculate the percentage of PGPubs having *identical* main symbol class codes, main classification symbols, and sets of classification symbols to their corresponding issued patents (*capture rate*)

Methodology: Analysis Methods

- For analysis 3, we further calculate the *Jaccard Coefficient*

$$-\mathrm{J} = \frac{|\{\mathrm{PGPub}\} \cap \{\mathrm{Patent}\}|}{|\{\mathrm{PGPub}\} \cup \{\mathrm{Patent}\}|}$$

  - {PGPub}: PGPub's set of classification symbols
  - {Patent}: Issued patent's set of classification symbols
- All 3 analyses also do not consider the hierarchical relationship among the symbols

Methodology: Analysis 1

PGPub no./Patent no.	PGPub symbols	Patent symbols
20140289912/8,955,161	**850/18**	**850/1**; 250/339.11; 250/339.14; 73/105; 850/5; 850/50; 850/6
20120124680/8,955,160	**726/34**	**726/34**
20110252484/8,955,159	**726/32**	**726/32**; 380/201; 705/57; 726/27; 726/31; 726/33

- All three pairs have identical main symbol class codes
  - The PGPub classification symbols have completely capture the content of the issued patens
  - Capture rate=100% (3/3)

Methodology: Analysis 2

PGPub no./Patent no.	PGPub symbols	Patent symbols
20140289912/8,955,161	**850/18**	**850/1**;250/339.11;250/339.14;73/105;850/5; 850/50;850/6
20120124680/8,955,160	**726/34**	**726/34**
20110252484/8,955,159	**726/32**	**726/32**;380/201;705/57;726/27;726/31;726/33

- 2 out of the three pairs have identical main symbols
  - The PGPub classification symbols have partially capture the content of the issued patens
  - Capture rate=67% (2/3)

Methodology: Analysis 3

PGPub no./Patent no.	PGPub symbols	Patent symbols
20140289912/8,955,161	**850/18**	**850/1**; 250/339.11; 250/339.14; 73/105; 850/5; 850/50; 850/6
20120124680/8,955,160	**726/34**	**726/34**
20110252484/8,955,159	**726/32**	**726/32**; 380/201; 705/57; 726/27; 726/31; 726/33

- Only 1 out of the three pairs have identical sets of symbols: 726/34
  - Capture rate=33% (1/3)
- Jaccard Coefficients are
  - 1st pair: 0
  - 2nd pair: 1
  - 3rd pair: 1/6

Result: For 235,000 pairs

- The capture rates from the 3 analyses

Analysis	Pairs	Capture rate
1: Comparing main classification class codes	183,024	77.89%
2: Comparing main classification symbols	85,584	36.42%
3: Comparing entire sets of classification symbols	14,958	6.37%

Result: For 235,000 pairs

- For Analysis 3, we further divide the 235,000 pairs of patent-PGPub pairs into 5 categories, and look at their average Jaccard Coefficients separately.

Category		Pairs	Share	Avg. J
1:	$\{PGPub\} = \{Patent\}$)	14,958	6.37%	1
2:	$\{PGPub\} \neq \{Patent\}$, $\{PGPub\} \cap \{Patent\} = \emptyset$	89,981	38.30%	0
3:	$\{PGPub\} \neq \{Patent\}$, $\{PGPub\} \subset \{Patent\}$	63,057	26.84%	0.34
4:	$\{PGPub\} \neq \{Patent\}$, $\{Patent\} \subset \{PGPub\}$	10,693	4.55%	0.45
5:	$\{PGPub\} \neq \{Patent\}$, $\{PGPub\} \not\subset \{Patent\}$ $\{Patent\} \not\subset \{PGPub\}$, $\{Patent\} \cap \{PGPub\} \neq \emptyset$	56,277	23.95%	0.22

Conclusion

- It seems that **the best practice for applying PCA on PGPubs** using a commercial analytic system is by Method 1 (i.e., **counting the PGPub main symbol class codes**)
  - For Methods 2 and 3 (i.e., counting the main classification symbols and counting all classification symbols), the analytic result such would either miss or carry too much noise to be trustworthy.

Conclusion: Some class codes are trustworthy and some don't

Top ten class codes			Bottom ten class codes		
Class	No. of assignments	Capture rate	Class	No. of assignments	Capture rate
385	711	95.64%	703	769	65.93%
320	681	93.10%	426	771	65.37%
716	791	93.05%	204	619	64.46%
347	2,147	92.92%	700	1,734	61.65%
343	678	91.89%	264	800	60.62%
701	2,938	91.42%	424	4,846	60.30%
381	1959	91.37%	137	726	57.85%
365	2,496	91.27%	524	982	55.30%
473	686	91.11%	427	1,168	55.14%
219	750	90.67%	428	3,056	54.58%

Thank you