Modeling of Chemical–Mechanical Polishing: A Review

Gerd Nanz and Lawrence E. Camilletti

***Abstract*— This paper gives a survey of the status of today's modeling of chemical–mechanical polishing (CMP). Most existing models describe specific aspects of CMP, such as the flow of the slurry or the bending of the polishing pad. However, as yet no model describes the entire available process. This paper critically reviews existing models with respect to generality. In particular, the different assumptions of the models are investigated. Furthermore, the models are compared and the controversial treatment of physical effects is discussed.**

I. Introduction

A NOVEL technique for planarization is of growing interest, since conventional planarizing methods such as flowing oxide layers do not give the required global planarity for advanced processes. Multilevel interconnects and the use of 3-D packaging require sophisticated methods to planarize the surfaces of wafers for subsequent device processing. For five or more layers of a logic device at least one layer should be perfectly planar [14]. Lack of planarity may lead to severe problems for lithography (insufficient focus depth) and dry etching in sub 0.5-μm IC processes [7].

Several methods are known to achieve a higher level of planarization: (chemical–mechanical) polishing, laser reflow, coating with spin-on glasses, polymers, and resists, thermally reflowing materials, dielectric deposition [14], and flowable oxides.

In the following sections the present status of modeling CMP available from the literature is discussed. First, the basic ideas of CMP are described. Then an overview of existing models for CMP is given, including a discussion of their capabilities. Finally, the models are compared pointing out some controversial approaches to describe physical effects.

A. CMP

A schematic of a CMP machine is shown in Fig. 1 (view from top and cross section). A CMP machine uses orbital, circular and lapping motions. The wafer is held on a rotating carrier (holder) while the face being polished is pressed against a resilient polishing pad attached to a rotating platen disk. For oxide or silicon polishing, an alkaline slurry of colloidal silica (a suspension of SiO_2 particles) is used as the chemical abrasive. The size of the particles varies in literature between 100 Å [19] and 3 μm [2]. According to [15], the size of the particles is between 600 Å and 800 Å forming agglomerates of a size of 2500 Å in diameter. The slurry is carried to the wafer by the porosity of the polishing pad. This slurry chemically attacks the wafer surface, converting the silicon top layer to a hydroxilated form (with the OH^- radical) which is more easily removed by the mechanical abrasive. The details of the formation of this top layer are not yet well understood [8], [23].

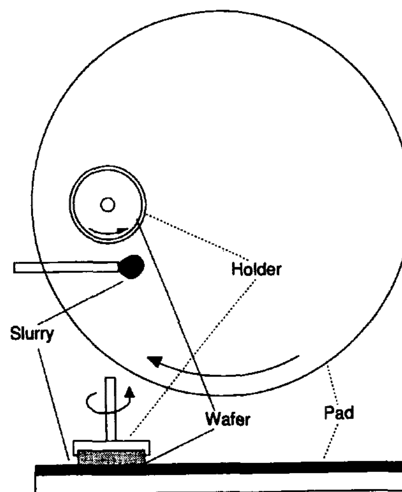

Fig. 1. Schematic of chemical–mechanical polishing technique.

Gross mechanical damage of the surface is prevented by the fact that the colloidal silica particles in the slurry are not harder than the oxide being removed [8]. Otherwise the quality of the surface planarity would be limited by the diameter of the silica particles.

CMP needs fewer steps compared to deposition/etchback [6]. Furthermore, CMP uses nontoxic substances, has a good removal selectivity, and a good rate control. Typical values of some essential parameters in the CMP process are given in Table I.

Another advantage of CMP lies in the global planarization. Since the sizes of flat areas on a chip become smaller the quality of local planarization (in a global sense) becomes worse. Additional difficulties may arise for the filling of small holes. CMP reduces defect density, according to [15]. Shorts/defects due to residual metal can be significantly reduced, as reported in [4]. However, real-life processes, if not controlled properly, will add a significant number of defects. For instance, as the pressure (down force) is increased, the

Manuscript received December 10, 1994; revised March 23, 1995.
G. Nanz is with Digital Equipment Corporation, Favoritenstrasse 7 (CEC Vienna), A-1040 Vienna, Austria.
L. E. Camilletti is with Digital Equipment Corporation, Hudson, MA 01749 USA.
IEEE Log Number 9414529.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

TABLE I
TYPICAL VALUES FOR CMP PARAMETERS

Quantity	Reference	value
Removal of SiO_2 (thermal)	[6]	60nm/min to 80nm/min
Removal of SiO_2 (LPCVD)	[6]	100nm/min to 150nm/min
Polishing time	[9]	10min
Pressure (pad/wafer)	[9]	6psi
Velocity of pad	[9]	10rpm
Velocity of wafer	[9]	12rpm

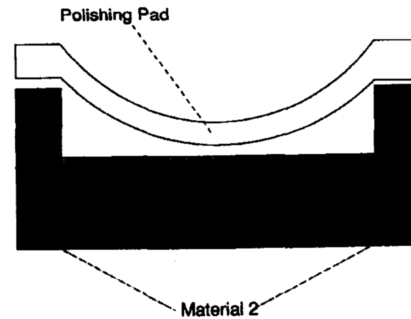

Fig. 2. Pad bending during CMP.

risks also increase. There is a greater propensity for dislocation and defect generation and the possibility of diffusion or penetration of the slurry contaminants below the surface exists.

CMP is applied for several types of structures: Bare silicon before any processing starts, metallization, the intermetal layer dielectric (ILD), and process silicon with dielectrics and metals. In the latter two cases the selective removal of materials such as aluminum, tungsten, and SiO_2 is necessary.

II. MODELING AND SIMULATION

The main feature of CMP, namely the removal of the material, is described by Preston's equation:

$$\frac{dT}{dt} = K \cdot \frac{N}{A} \cdot \frac{ds}{dt} \tag{1}$$

where T denotes the thickness of the wafer, N/A denotes the pressure caused by the normal force N on the area A, s is the total distance traveled by the wafer, and t denotes the elapsed time. This means that the removal of the wafer material is proportional to the pressure and the velocity of the rotation. Any physical considerations are put into Preston's constant K, which often is considered a proportionality constant (independent of pressure and velocity), but may also contain advanced physics and include the effects caused by the chemical reactions. However, the removal rate tends to dominate in real-life situations.

Some models recently published are described below and discussed with respect to their limitations and their range of applicability.

A. Model by Sivaram

Sivaram *et al.* [16], [17] have developed a model which describes the removal according to (1) taking also into account the bending of the polishing pad.

As an important physical effect CMP has to deal with bending [4]. A schematic of the pad and the wafer is drawn in Fig. 2. Due to the pressure on top of the polishing pad, the pad behaves locally like a beam which is supported by the blocks of material 2. Assuming that material 2 is inelastic, the deflection of the pad can be easily calculated. The solution v of the differential equation (2) describes the deflection of a beam with length l

$$\frac{d^4 v}{\cdots}$$

where E and I are Young's modulus (elasticity modulus) and the moment of inertia, respectively, and $w(x)$ denotes the load on the beam.

For a uniform load $w(x) = w_0$ and for the boundary conditions $v(x) = 0$ and $M(x) = 0$ for $x \in \{0, l\}$, where M denotes the momentum (second derivative of $v(x)$), one obtains (3) as the solution of (2)

$$v(x) = \frac{w_0}{24 \cdot E \cdot I} \cdot (x^4 - 2 \cdot l \cdot x^3 + l^3 \cdot x). \tag{3}$$

Due to the deflection of the pad it may occur that material 1 is affected by the pad. Therefore (3) is also a measure for the planarity which can be achieved.

Discussion: The treatment of the deflection of the polishing pad is important since the quality of the planarization is affected by the bending. However, Sivaram *et al.* consider only the effects of the bending between two neighboring peaks on the wafer. For a rigorous treatment of the deflection of the pad a wider range of the wafer must be considered requiring the calculation of the deflection of a multiply-supported beam. Additionally, the model assumes that the product $E \cdot I$ for the polishing pad is the same or higher than for the platen (holding the polishing pad). This is usually not true. If the polishing pad is softer than the platen, the pad is compressed while the platen stays nearly flat. This effect is not included in the approach for the calculation of the beam bending.

This model deals only with two-dimensional cross sections of the structure and neglects the slurry flow—thus the application is limited. For modeling purposes of the entire CMP process, a completely three-dimensional simulation would be necessary.

B. Model by Burke

A model dealing with the polishing rate depending on the degree of nonplanarity has been proposed by Burke [1]. The model has two stages: An analytical model which is based on the closed solution of a simple ordinary differential equation and a more complex model which iteratively adapts the polishing rate to the actual nonplanarity.

D_0 denotes the percent polish rate of areas which are low compared to the polishing rate for the blanket-wafer ("down" polishing rate). This means that for small D_0 planarization is good (ideal for 0), and planarization stops for D_0 close to 1, since then the lower areas are polished in the same

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

height associated with D_0. Burke observes that for $D_0 > 0.3$ the 'down' polishing rate is linear and for $D_0 < 0.3$ it is logarithmic.

For $D_0 < 0.3$, and a constant polishing rate U for 'up' areas the polishing rate D for 'down' areas, is given by (4), where S is the actual step height, i.e., the amount of nonplanarity

$$D = \left(1 - (1 - D_0) \cdot \frac{S}{S_0}\right) \cdot U. \quad (4)$$

The governing differential (5) looks similar to Preston's equation (1), but it is different

$$-\frac{dS}{dt} = \frac{1 - D_0}{S_0} \cdot S \cdot U \quad (5)$$

where U has the dimension of a velocity similar to the term ds/dt on the right side of (1). However, U is a removal rate and not the relative velocity between the pad and the wafer.

Burke integrates (5) and obtains the solution (6)

$$\frac{S}{S_0} = \exp\left(-\frac{U \cdot t \cdot (1 - D_0)}{S_0}\right). \quad (6)$$

Burke also proposes a more advanced differential model. The surface of the wafer is represented in the computer program as a two-dimensional topography given by the user. Then the polishing rate is adapted at each point according to the neighboring points similarly to the analytical model. Burke claims that corner rounding is predicted with sufficient accuracy.

Discussion: This model takes into account the type of the nonplanarity of the wafer surface and adjusts the polishing rate accordingly. The model does not deal with pad bending, asperities or the fluid flow. Furthermore, the model is empirical and does not address the dependence of the polishing rate on the pressure. Even though it takes into account the entire wafer (and not only two-dimensional cross sections) it covers only a small part of the entire CMP process.

C. Model by Warnock

A model which allows the quantitative analysis of the absolute and the relative polish rate for different sizes and pattern factors has been presented by Warnock [22]. It is a microscopic mathematical model which is completely phenomenological. Due to the nonlocal nature of CMP, different length scales are necessary which are determined by the flexibility and hardness of the polishing pad. The horizontal length scale is determined by the pad deformation and the vertical length scale by the pad roughness. In this model, the polishing rate P_i at each point i of the wafer is given by (7)

$$P_i = \frac{K_i \cdot A_i}{S_i} \geq 1 \quad (7)$$

where K_i is the kinetic factor (horizontal component), A_i is the accelerating factor (higher points on the wafer), and S_i is the shading factor (lower points on the wafer). A schematic surface demonstrating the meaning of the coefficients K_i, A_i, and S_i is shown in Fig. 3. In lower regions S_i is large, thus reducing P_i; in higher regions A_i is large, thus increasing P_i.

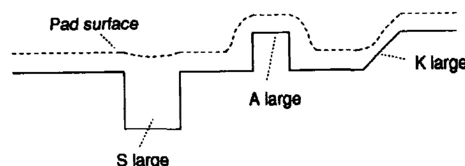

Fig. 3. Arbitrary surface with effects on the factors of the polishing rate.

There is also a direct reciprocity between the A_i and S_i at all n points, which is given by (8). It includes the assumption that the polishing rate is linearly proportional to the pressure

$$\sum_{i=1}^{n} \frac{A_i}{S_i} = n \quad (8)$$

$$S_i = \exp\left(\frac{\overline{\Delta z_i}}{z_0}\right). \quad (9)$$

S_i is given by (9), where z_0 is a scaling factor for the vertical length scale and $\overline{\Delta z_i}$ describes how much the surrounding topography protrudes above point i. The exponential function has been chosen to account for the rough nature of the pad. However, Warnock agrees that an improved formulation could be found, even though he claims that the model is not very sensitive with respect to the formulation

$$\overline{\Delta z_i} = \frac{1}{2 \cdot \pi} \cdot \int_0^{2\cdot\pi} \int_0^{\infty} z(r, \theta) \cdot W(r)\, dr\, d\theta \quad (10)$$

$$\overline{\Delta z_i} \approx \frac{1}{2 \cdot \pi} \cdot \int_0^{2\cdot\pi} z(r_m, \theta) \cdot W'(r_m)\, d\theta \quad (11)$$

$$W'(r_m) = \frac{1}{\cosh\left(\frac{r}{r_0}\right)}. \quad (12)$$

$\overline{\Delta z_i}$ is obtained by integration over the surrounding topography of point i. In (10), $z(r, \theta)$ is the vertical height at the coordinates r and θ with respect to point i, and $W(r)$ is a weighting function describing the horizontal length scale over which the pad deforms. Under reasonable assumptions (10) can be simplified to (11), where r_m is the value of r which maximizes $W(r)$ (as a function of θ). $W'(r_m)$ as chosen by Warnock is given in (12). Thus, the deformation length scale is r_0.

The A_i can be determined in an iterative process from the S_i. Warnock claims that uniqueness is given and that a solution is obtained 'after a large enough number of iterations.'

The K_i are determined by calculating an effective vertical component of the horizontal polish rate. Thus, $K_i = 1 + K_0 \cdot \tan \alpha_i$, where K_0 is a model parameter, and α_i is the local angle between the horizontal and the surface.

Discussion: This model gives a reasonable approach for defining the dependence of the polish rate on the wafer shape, even though it is completely phenomenological. In particular this model takes into account all geometrical cases; therefore, it might be general enough to cover a part of the simulation of the entire CMP process. Furthermore, it is not limited to

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

D. Model by Yu

The model by Yu *et al.* [23] deals with the dependence of the removal rate on the asperity of the polishing pad. From measurements asperities can be observed all over the polishing pad. The surface height variation is reported to be more than 100 μm on a 200 μm × 200 μm polishing pad. The asperities are assumed to be spherical at the summit and that the variations in height and radius are Gaussian-distributed in the model. Then Preston's constant K in (1) can be modified according to the new model, splitting it into three parts: One is a constant only determined by the pad roughness and the elasticity, one is a factor of surface chemistry and abrasion effects, and one is related to the contact area thus accounting for the asperity [15], i.e., the contact properties of the asperities depend on the width of a trench.

The behavior of the pad is assumed to be viscoelastic in the model, which means that the deformation of the pad can be taken into account.

Discussion: The model gives an approach how to deal with the asperities of the pad. However, it is not clear whether or how the asperities affect the global quality of planarization. A global planarization quantity of 200 Å over a distance of 0.5 cm has been reported [20]. This makes further investigations necessary as to how this approach fits into a general CMP simulation model. Additionally, the validity of the assumptions for the radius and the height of the asperities seem to need further research.

The reported variation in height of the polishing pad certainly overshadows the nonplanarities of the wafer. It is even comparable with the thickness of the fluid layer between the pad and the wafer. This question is neither treated by the model nor discussed.

E. Models by Runnels

Runnels *et al.* [10]–[12] propose several models accounting for the stress in the polishing pad and the fluid flow as well as the removal of material by erosion.

1) Flow of the Slurry: Runnels *et al.* propose a model accounting for the fluid flow between the wafer and the pad [11].

A wafer of radius 10 cm and spherical curvature rotates about its axis of symmetry, which is approximately 30 cm from the pad's rotational axis. The wafer glides at an angle of attack θ upon the slurry film whose thickness is denoted by h. The wafer carrier is mounted on a gimbal mechanism to prevent the wafer snagging on the pad. The model focuses on the flow of the slurry between the wafer and the pad. The flow simulation is embedded in an iterative scheme for determining h and θ. For the flow simulation, it is assumed that the wafer and the pad are rigid and smooth. Therefore, both the pad and the wafer can be described by boundary conditions for the flow of the slurry. Even though the slurry contains particles (with a magnitude of approximately 1 μm) the flow of the slurry is assumed to be Newtonian with a constant viscosity. Thus, the flow is given by the steady-state incompressible Navier–Stokes equations in three space dimensions (13)

$$\vec{u} \cdot \nabla \vec{u} = -\frac{1}{\rho} \cdot \nabla p + \frac{\mu}{\rho} \cdot \nabla^2 \vec{u}$$

where ρ denotes the density, μ is the dynamic viscosity, p is the pressure, and $\vec{u}$ is the vector-valued function of the velocity at any point in the flow.

The simulation domain consists of a thin disk bounded by the surfaces of the wafer and the pad, respectively, and a ring, representing the flow around the outside of the pad. For the wafer, the pad, and the sidewalls of the ring no-slip boundary conditions are applied—thus giving the fluid the velocity of the wafer, the pad, and the flow in the ring. For the remaining surfaces (mainly of the ring) stress-free boundary conditions are applied, thus allowing the fluid to enter and leave freely.

Two conditions have to be fulfilled for the determination of the fluid layer. First the fluid layer must support the wafer carrier, including the applied load during polishing. The force F on the wafer surface from the fluid flow is given by (14), where σ is the stress tensor related to the flow field by (15). δ_{ij} denotes the Kronecker symbol

$$F = \int_{\text{wafer surface}} \sigma \cdot \vec{n}\, dA \tag{14}$$

$$\sigma_{ij} = -p \cdot \delta_{ij} + \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right). \tag{15}$$

The second condition is that the fluid is stable. This requires the moment of the force caused by the fluid film to have components which vanish in the plane perpendicular to the carrier's axis of rotation. The moment M_f about the gimbal point from the fluid flow is given by (16), where $\vec{R}_g$ denotes the distance of each point on the wafer measured from the gimbal point

$$M_f = \int_{\text{wafer surface}} \vec{R}_g \times \sigma \cdot \vec{n}\, dA. \tag{16}$$

The thickness of the fluid layer h and the angle of attack θ are calculated by an iterative scheme satisfying that M_f vanishes.

Discussion: This model analyzes fluid flow. It turns out that stringent assumptions are necessary to put this complex problem into a mathematical model. The wafer surface is assumed to be spherical with a large radius in the model, which means that all questions about the polishing mechanism and the structure of the wafer surface are neglected. Furthermore, Runnels admits that the validity of their assumptions is not completely clear and that their model can give only a qualitative estimation of the fluid layer thickness.

The main result of this model is the thickness of the fluid layer between the wafer and pad. The simulation must be made for the entire wafer and be posed as a fully three-dimensional problem in order to determine this thickness correctly. The solution of the Navier–Stokes equations can be used for the modeling of the material removal rate.

To the authors' knowledge this is the first work dealing with the slurry flow, which is an essential factor in modeling CMP.

2) Removal by Erosion: Runnels proposes a model which calculates the removal as a consequence of erosion due to the slurry flow [10].

This two-dimensional feature scale model, which takes into account cross sections of the CMP machine, is mainly

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

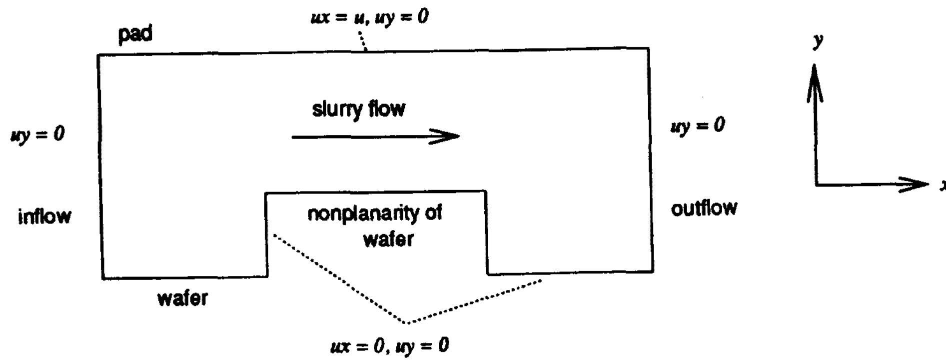

Fig. 4. Geometry of erosion simulation domain.

The fluid flow around a nonplanarity on the wafer (see Fig. 4) is given by the solution of the two-dimensional Navier–Stokes equations (13). The boundary conditions are also given in Fig. 4, where ux and uy denote the unknown velocities in x- and y-direction, and u is the velocity of the pad. Again, the surface stress tensor is calculated by (15). The erosion rate V_n in the normal direction is given by (17) as a function f of the time-dependent tangential and normal stresses σ_t and σ_n

$$V_n = f(\sigma_t(t), \sigma_n(t)). \tag{17}$$

Equation (17) is integrated in time and gives the shape of the wafer surface for any time.

Runnels relates V_n to Preston's equation by an approximation following heuristic arguments with formulas (18)–(22). The constant C represents microfracturing and chemical aspects. A is the area of the wafer and P is the bulk pressure. An estimation of V_n is given in (18). Equation (19) together with the approximation for slider bearings (20) leads to the formulation (21) which can be rewritten by (22), which is the same as Preston's equation (1)

$$V_n = C \cdot \sigma_t^2 \tag{18}$$
$$\sigma_t \approx \mu \cdot \frac{u}{h} \tag{19}$$
$$h \propto \sqrt{\frac{\mu \cdot u}{P \cdot A}} \tag{20}$$
$$V_n \propto C \cdot \left(\mu \cdot \frac{u}{\sqrt{\frac{\mu \cdot u}{P \cdot A}}} \right)^2 \tag{21}$$
$$V_n \propto \mu \cdot A \cdot u \cdot P. \tag{22}$$

Runnels also includes the stress dependence of the chemical reaction and an approximation for the rotation of the pad. This increases the complexity of the expression for V_n and is still a matter of research and further modeling work.

Discussion: The model is mainly based on the solution of the two-dimensional Navier–Stokes equations thus simulating the flow of the slurry in a cross section. The simulation setup model. Since the Reynold's number seems to be small for this type of application no major numerical problems should be encountered. Therefore, the model is a good base for further extensions such as accounting for the pad deflection. Runnels leaves several questions about the physical modeling of the erosion law open and uses completely heuristic arguments to fill gaps in his model. The influence of the fluid flow on the erosion rate is only given by the stress tensor.

The question of the boundary conditions for the Navier–Stokes equations along the wafer surface (vanishing velocities in both coordinate directions) should be reconsidered.

3) Deformation of Pad: Runnels *et al.* propose a model which accounts for the deflection of the polishing pad at the wafer edges and the resulting stress distribution [12].

Preston's equation (1) is reformulated leading to (23), where R is the removal rate

$$R = K \cdot P \cdot ||v||. \tag{23}$$

Runnels *et al.* claim that $||v||$ should be replaced by $||\sigma||$, where σ is the vector-valued shearing stress acting in the plane of the wafer surface while P can be interpreted as a stress perpendicular to the plane of the wafer.

Several assumptions are made: 1) The transfer of stresses between wafer and pad are neglected, 2) the pad is assumed to be elastic, even though it is known to be viscoelastic—Runnels *et al.* claim that for the typical speeds of rotation an elastic representation is suitable, 3) the slurry flow is neglected. Therefore, two extremes are considered: Once the pad and the wafer adhere to each other, and once they are allowed to slide freely without stresses.

The boundary of a planar wafer is considered for the analysis. The governing equations for the axisymmetric case (in polar coordinates) are given by (24), where σ_r and σ_z are the normal stresses in the radial and vertical directions, respectively, and τ_{rz} is the shear stress

$$\frac{\partial \sigma_r}{\partial r} + \frac{\partial \tau_{rz}}{\partial z} + \frac{\sigma_r - \sigma_\theta}{r} = 0$$
$$\frac{\partial \tau_{rz}}{\partial r} + \frac{\partial \sigma_z}{\partial z} + \frac{\tau_{rz}}{r} = 0. \tag{24}$$

The deflection of the pad is then related to the stress through

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.