ISSN: 2688-9323

Textual Defense under Contextual Pressure: R&D Accounting Choices, Regulatory Inquiries, and the Evolution of Footnote Obfuscation

Shijia Chen

University of Shanghai for Science and Technology, Shanghai, China

Abstract

With the increasing stringency of financial statement presentation formats in China's capital market and the normalization of "penetrative" inquiry supervision, corporate earnings management is gradually shifting from numerical to covert textual manipulation. This paper explores how the management of STAR Market semiconductor enterprises, operating under high technical cognitive barriers, utilizes audited financial statement notes to implement text obfuscation strategies. Based on Eisenhardt's multi-case theory construction paradigm, this study selects four chip enterprises with distinctly polarized characteristics and uses NLP technology to conduct longitudinal tracking and cross-case comparisons of their R&D disclosure texts from 2019 to 2025. The findings reveal that deep financial pressure is more likely to trigger "syntactic obfuscation" than aggressive accounting policy choices, as pressured enterprises build implicit defenses by stretching syntactic complexity. Furthermore, rather than deliberately piling up professional terms, management expands text using template-like, redundant descriptions lacking incremental information. This creates a "watering-down effect" that reduces the concentration of core accounting terms under the guise of formal compliance, thereby increasing readers' cognitive load. Additionally, external inquiry supervision exerts a significant "asymmetric impact" on heterogeneous enterprises. While it prompts stable firms to clarify information, it inadvertently intensifies the defensive instincts of problem enterprises, driving them into a strategic loop where more regulation leads to more obfuscation. Ultimately, this study breaks through the limitations of traditional MD&A tone analysis and static word frequency statistics. It enriches management obfuscation theory within specific institutional contexts and provides a forward-looking early warning scale, enabling regulatory authorities to shift from single-content compliance reviews to the "dual monitoring of content and structure.

Keywords

Textual Obfuscation; Notes to Financial Statements; Regulatory Inquiry Letters; Natural Language Processing (NLP).

1. Introduction

During China's economic transformation period centered on self-reliance and strength in science and technology, high-tech enterprises such as semiconductor companies quickly listed on the Sci-Tech Innovation Board (STAR Market) by taking advantage of the policy benefits of the registration system. They have become the core engine driving the high-quality development of the capital market. In order to further enhance the market transparency of the research and development innovation information of these enterprises, the Ministry of Finance of China implemented a reform of financial statement formats in 2018, mandating that research and development expenses be moved from the concealed footnote items to be separately listed in the profit statement [1]. This significant institutional change, aimed at squeezing the space

ISSN: 2688-9323

for digital profit management through the independent listing of research and development expenses in the profit statement, has quietly shifted the game between management and the market to a more concealed textual domain. Facing extremely high technical barriers and increasingly strict exchange inquiries and supervision, enterprises are confronted with unprecedented tensions between responding to market concerns, meeting compliance requirements, and concealing potential operational risks. This has led the financial statement footnotes, which were originally an objective supplementary explanation, to gradually evolve into a new battlefield for management to implement implicit risk concealment and text obfuscation [2].

Looking at the academic exploration process of enterprise information disclosure quality, the early research mostly focused on soft and intuitive texts such as Management Discussion and Analysis (MD&A), emphasizing the manipulation of tone and impression management [3]. Even with the comprehensive penetration of natural language processing (NLP) technology, the academic community has gradually established the cognitive connection between text readability and earnings management. However, the existing research landscape, mostly delineated by the empirical logic of developed Western markets, still shows a significant lack of specificity when facing the unique institutional context of China's Sci-Tech Innovation Board, which has "high proportion of retail investors"[4] and "strong inquiry constraints" [5]. It is difficult to objectively present the real stress state of local enterprises under the distinctive regulatory environment. Compared to the clearly intentional MD&A, the audited financial statement notes, as the theoretical "hard information" carrier, are often regarded as data interpretation without any impurities by the outside world [3]. However, this preconceived rigorous image has instead relegated its internal text manipulation logic to the periphery of academic observation, forming a cognitive "black hole" [3]. If we further examine the research paradigm, traditional large-sample empirical studies often adhere to static word frequency statistics, simply equating "text confusion" with the rigid accumulation of technical terms, without fully recognizing the means by which management uses boilerplate and other methods to dilute information [6]. This limitation at the measurement level precisely hinders our ability to dynamically capture, in actual research observations, how the underlying disclosure strategies of enterprises evolve into a set of micro and intricate stress logic in the moments when they are deeply trapped in financial pressure or facing regulatory shocks.

Based on the identified theoretical gaps and practical challenges, this study aims to delve beyond the cold financial figures and thoroughly analyze the actual disclosure logic of semiconductor enterprises under different financial backgrounds and regulatory circumstances. Specifically, this paper intends to explore how, when management adopts an aggressive policy of capitalizing research and development or faces the heavy pressure of consecutive years of huge losses in their financial performance, they implement implicit defense and text obfuscation through skillfully adjusting "sentence complexity" and "text redundancy" [7]. At the same time, it attempts to reveal whether the external intervention mechanism of exchange inquiries can effectively discipline the enterprises' improper disclosures or, in certain situations, becomes a catalyst for higher-level text obfuscation [2].

By leveraging NLP technology, after conducting a long-term longitudinal tracking and multi-dimensional horizontal comparison of the R&D notes of these four enterprises from 2019 to 2025, the data not only conclusively confirmed that the deep financial pressure was the underlying logic that triggered syntactic confusion [8], but also revealed, within the lengthy text, a "watering-down dilution effect" hidden beneath the lengthy content. Under the real review pressure, those enterprises under pressure who were eager to conceal risks did not, as expected by traditional theories, deliberately pile up obscure words or Jargon. Instead, under the guise of precise compliance, they tended to inject a large amount of non-informative disclosures and boilerplate and verbosity, thereby diluting the absolute concentration of core

ISSN: 2688-9323

accounting characteristic words and quietly increasing the cognitive load of external investors. This defensive instinct is particularly evident when facing inquiries from exchanges - external regulation inevitably causes significant asymmetric impacts on heterogeneous enterprises: it can certainly force stable enterprises to further clarify the facts. But at the same time, it inadvertently triggers the sensitive nerves of problem enterprises, stimulating them to accelerate the completion of the strategic loop from passive disclosure to substantive confusion.

This meticulous dissection of "syntactic stretching" and "lengthy padding" in parallel precisely constitutes the most core marginal contribution of this study. It completely breaks away from the previous path dependence on MD&A tone analysis or mechanical word frequency statistics [6], and re-anchors the scrutiny on the "audited financial notes" in the Chinese capital market, which have a higher signal-to-noise ratio and possess legal validity. This not only greatly expands the theoretical depth of the management confusion theory in the local context, providing a highly explanatory theoretical anchor for the frequent research and disclosure anomalies of R&D information under the registration system. At the same time, from the regulatory practice perspective, it also provides a highly practical and predictive policy benchmark for policy makers to make the transition from static single-content compliance review to "dual monitoring of content and structure" in digital and intelligent governance, which is conducive to policy formulation.

2. Institutional Background and Theoretical Evolution

2.1. Changes in Chinese Standards and Market Characteristics

In the "Notice of the Ministry of Finance on Revising and Issuing the 2018 General Enterprise Financial Statement Format" released by the Ministry of Finance on June 15, 2018, a new round of revisions was made to the financial statement format of listed companies. The focus was on adding a "Research and Development Expenses" line item in the profit statement, mandating that listed companies change the previously recorded research and development expenses as a detail item under "Administrative Expenses" in the notes to the financial statements, and instead report them separately in the profit statement. This reform of the reporting format has further significantly standardized the presentation of financial statements, thereby improving the quality of accounting information in the annual reports of listed companies and reducing information asymmetry caused by the high uncertainty of research and development activities [1]. Existing studies have pointed out that this "on-sheet" presentation method enhances the salience of research and development information, making the previously overlooked detailed items in the notes more "attractive", thus helping capital market participants more efficiently collect, transmit, and compare the research and development innovation investments of different companies.

However, the standardization of reporting formats has also triggered a deep game of disclosure behavior by management. On one hand, mandatory independent reporting of research and development expenditures may increase the proprietary costs of enterprises [9]. More transparent information disclosure may allow competitors to peek into the core business secrets of enterprises through financial reports, thereby damaging the competitive advantages of enterprises. On the other hand, when the reporting rules at the numerical level become stricter and more difficult to circumvent, enterprises with motives for earnings management or the need for concealment often turn to more covert textual manipulation strategies. In the Chinese capital market, retail investors dominate [4]. However, due to the extremely complex underlying technologies involved in semiconductor design, such as SoC, EDA, and FinFET, retail investors often have significant "information blind spots" when facing highly specialized research disclosure. This extremely high technical cognitive barrier further amplifies the

ISSN: 2688-9323

operability and motivation of management to implement information asymmetry through textual disclosure [8].

2.2. Evolution of the Regulatory Environment: The "Penetrative" Deterrence of Inquiry Letters

In recent years, the "annual report inquiry letter" system centered around exchanges has become an important part of China's distinctive securities regulatory framework. This non-punitive regulatory tool aims to enhance the transparency of financial reports through public inquiries and responses [5]. Unlike traditional compliance checks, inquiry letters exert "penetrating" regulatory effects by deeply questioning specific accounting estimates (such as the rate of capitalization of research and development) [10]. Moreover, inquiry letters have a significant governance effect [5]. When companies face regulatory doubts, management must provide detailed explanations of the disputed matters in subsequent annual reports or response letters. However, this regulatory deterrence may also have unintended consequences: to meet compliance requirements in form while hiding core risks, management may adopt defensive disclosure strategies, increasing redundant information to distract regulators' attention [11]. This dynamic regulatory game makes the quality changes in the disclosure texts of companies an important window for observing the true financial status of enterprises [12].

2.3. Theoretical Development: From "Impression Management" to "Text Confusion"

The early research on the analysis of disclosure texts usually started from the MD&A section, mainly based on the theories of "impression management" and "signal transmission", suggesting that the management would adjust the tone of the text to influence investors' perception. Under this theoretical logic, high-quality enterprises tend to convey positive signals through clear and positive language to reduce information asymmetry and financing costs; while low-quality enterprises attempt to imitate this behavior [13].

However, as the regulatory environment has gradually become stricter and more comprehensive, the risk of blatant false statements has increased, and the management has turned to a more covert "text obfuscation" strategy. The management text obfuscation hypothesis suggests that when an enterprise has poor performance, illegal behavior, or financial pressure, the management will deliberately reduce the readability of financial reports and footnotes to conceal negative news [8].

This strategy mainly involves two operational methods: One approach is that the management may, in the explanatory text, lengthen sentences and pile up complex sentence structures to reduce the readability of the disclosed content, thereby meeting the "independent listing within the table" requirements in form while increasing the cost for investors to penetrate the financial fog and process the information [14].

The other method is that the management adds a large amount of insignificant boilerplate or verbosity to drown the core negative information in the sea of information, even causing investors to give up reading. This "dilution effect via length expansion" behavior results in a significant increase in the text length, but the substantive information density is significantly diluted due to the expansion of the denominator. This confusion method based on language structure is more difficult to be detected by traditional detection tools than simple tone manipulation, and it also increases the difficulty for external audits and investors to conduct in-depth analysis, becoming the focus of current multiple case studies [3].

ISSN: 2688-9323

3. Research Design and Sample Description

3.1. The Applicability of Eisenhardt's Multi-Case Theory Building Method

The dynamic evolution of management's disclosure strategy is not an isolated financial phenomenon; it is deeply rooted in the complex interaction between the firm's earnings management motives and external regulatory shocks, such as inquiries from securities regulatory authorities [15]. To penetrate the opaque textual surface and uncover the underlying triggering mechanisms, this paper employs the multi-case theory construction method advocated by Eisenhardt (The "Eisenhardt Method") for its research [16]. The core objective and advantage of this approach lie in "theory building", which can deeply address exploratory questions that lack obvious answers, such as "how" a specific phenomenon occurs [16]. Especially when the theory is not yet fully mature or needs to be extended for specific contexts - such as the unique R&D reporting reform after the establishment of China's STAR Market - multi-case studies, through the constant comparison of data and theory, can generate theoretical propositions that are more logically robust and have clearer boundary conditions than those from single-case studies [16]. In the research on the disclosure strategies of management, traditional cross-sectional large sample regressions often fail to capture the subtle mutations in the time dimension. In contrast, multi-case studies can achieve a deep integration of theoretical foundations and factual realities by depicting the game trajectory between enterprises and regulators in the time dimension [16]. Furthermore, this inductive research, through multiple perspectives of "within-case analysis" and "cross-case comparison", can identify anomalies emerging from quantitative data and further summarize them into qualitative theoretical models supported by underlying theoretical mechanisms [16].

3.2. Theoretical Sampling

In the specific case selection process, this article strictly adhered to the principle of theoretical sampling rather than statistical sampling, aiming to select typical samples that could reflect the polar types or matched pairs, in order to enhance the validity of the theoretical derivation [16]. The four semiconductor companies listed on the Sci-Tech Innovation Board selected for this study - VeriSilicon, Cambricon, SG Micro, and Montage Technology - together form a precise multiple comparison matrix.

Firstly, VeriSilicon, as the only sample in this study that has a significant capitalized research and development balance and has received regulatory inquiries, represents the "dual aggressive and shock-type" case. Due to the necessity of projects such as AIGC and smart transportation Chiplet solutions, the fluctuation in export revenue ratio, and cash flow pressure during its share issuance to specific targets in 2024, it encountered in-depth inquiries from the Shanghai Stock Exchange. The dynamic fluctuations of its text best reflect the stress response and text reconstruction of the management under capitalization decision-making and external regulatory deterrence [17]. In contrast, Cambricon represents the "textual defensive type under performance pressure" sample. Due to the extreme pressure situation of consecutive years of huge losses, the company received inquiry letters multiple times in 2022 to 2025 regarding core survival issues such as revenue recognition of intelligent computing clusters and commercialization prospects of edge product lines. Its text strategy exhibits a deep defense feature based on "performance compensation" and concealment of negative news [8].

To further verify the asymmetric impact of regulatory shocks on heterogeneous enterprises, this paper introduces SG Micro as a "robust and transparent" control. The company underwent in-depth review by the Shenzhen Stock Exchange due to the 1.38 billion yuan large goodwill generated from the acquisition of Yutai Semiconductor and the rationality of valuation issues in 2020. Even facing such sharp business and financial detail inquiries, the text performance of SG Micro under the background of expensing accounting constitutes an important reference for

ISSN: 2688-9323

examining "whether inquiry letters necessarily induce confusion strategies" [14]. Finally, Montage Technology, as a company that has capitalized research and development expenses and has not been affected by regulation, established the "pure baseline" of the semiconductor industry in its natural state.

Through this cross-case replication logic and multiple comparisons, this paper can summarize regular text disclosure strategies from different dimensions of financial pressure and regulatory feedback, thereby constructing a more robust theoretical proposition.

3.3. Data Collection and NLP Measurement

The corpus for this study was constructed over the entire period from 2019 to 2025. To ensure the scientific nature of the analysis, a rigorous process of data purification was carried out during the data collection stage. Through in-depth review of the footnotes of annual financial reports and regulatory response letters, only the paragraphs directly related to core accounting processing, detailed statements of the reports, industry technical characteristics, and regulatory interactions were extracted. This approach effectively eliminated potential marketing language noise in the management discussion and analysis (MD&A).

In the quantitative part, this paper utilized the Python programming language and its open-source data science libraries (pandas, re) to convert the complex qualitative descriptions into quantifiable feature scales. To restore the four dispersed footnote texts into a complete disclosure context, the program automatically removed null values and merged the four major types of texts into a unified analysis corpus. In the natural language processing (NLP) stage, this paper employed the mainstream jieba Chinese word segmentation tool to segment and slice the combined total text, with the aim of extracting multi-dimensional unstructured text features [6]. To overcome the segmentation errors of general dictionaries in professional fields, this study consulted relevant accounting standards and industry research reports, and manually constructed a 13-word "accounting compliance dictionary" including "development stage", "capitalization", "included in current profit and loss", and a 26-word "semiconductor technical jargon dictionary" including "fabrication", "wafer", "SoC", "FinFET". Before the word segmentation was executed, the code was used to forcibly protect these professional terms from being wrongly segmented, thereby ensuring the absolute accuracy of subsequent word frequency statistics.

Drawing on the design of Bozanic et al. [15] and the latest empirical studies from both domestic and international sources, this paper abandons the subjective tone analysis and instead, comprehensively constructs the following four core text variables from the objective dimensions of "length characteristics", "syntactic complexity" and "vocabulary density of traits":

(1) Total Length of Text. This indicator is used to measure the degree of redundancy disclosed by management. A significant increase in the text length is often an important benchmark for evaluating the quality of disclosure [15]. To eliminate layout differences, this paper uses regular expressions to remove all newline characters, tab characters, and spaces (\n \t \s \r), and takes the length of the cleaned pure Chinese characters and valid punctuation as the measurement standard for the total length of the text.

(2) Text Readability: Average Sentence Length (Avg_Sentence_Length). The average sentence length is regarded as the most crucial proxy variable in the academic community for measuring "management text obfuscation" (Obfuscation). The longer the sentence, the more modifiers and subordinate clauses it contains, and the greater the information processing cost and cognitive load for external investors. Management often uses this to complicate financial reports in order to conceal negative information [8]. This paper uses the standard Chinese sentence-ending punctuation as the delimiter to count the total number of sentences, and the measurement formula is:

ISSN: 2688-9323

$$Avg\_Sentence\_Length = \frac{Total\_Length}{Total\_Sentences} \quad (1)$$

(3) Accounting Compliance Word Density (Accounting_Density). This indicator is used to measure the degree of reliance of the disclosure text on the formatting template and the "watering-down effect". If an enterprise adopts a prudent and compliant disclosure strategy, it usually manifests as a dense accumulation of concise and easily understandable common accounting terms; while aggressive enterprises often need to add a large amount of non-standardized explanatory "bogus" content, resulting in a significant dilution of characteristic words. Its formula is:

$$Accounting\_Density = \frac{Count_{Accounting}}{Total\_Length} \quad (2)$$

(4) Comprehensive Confusion Index (Obfuscation_Index). This index measures the overall concentration of complex professional jargon (such as "accounting professional compliance boilerplate" and "semiconductor technology barriers terms") in the text [2]. Its measurement formula is:

$$Obfuscation\_Index = \frac{Count_{Accounting} + Count_{Semicon}}{Total\_Length} \quad (3)$$

By comparing the changes in these density ratios before and after the inquiry, this study is able to identify whether the management has used non-substantive descriptions to "water down and dilute" the information, thereby deliberately reducing the true readability of the information under the guise of formal compliance, in order to conceal their manipulation behavior.

4. Case Analysis: Disclosure Strategies under Situational Pressure

4.1. Descriptive Statistics of the Sample

Table 1. Summary Table of Research and Development Disclosure Text Characteristics of Sample Enterprises (Average Values for 2019 - 2025)

Firm Code	Firm Name	Mean R&D Capitalization Rate	Mean Footnote Length	Mean Sentence Length	Mean Accounting Word Density	Mean Jargon Density	Mean Obfuscation Index
688521	VeriSilicon	5.35%	12605.429	59.570	1.452%	0.144%	1.710%
688256	Cambricon	0.00%	8498.571	65.891	1.335%	0.125%	1.703%
300661	SG Micro	0.00%	10714.286	52.670	1.838%	0.000%	1.951%
688008	Montage Technology	0.00%	6137.857	55.184	1.890%	0.022%	2.139%

Based on the research methods and data processing approaches presented in Chapter 3, I first conducted a benchmark measurement of the characteristics of the R&D disclosure texts of four sample enterprises from 2019 to 2025. Through the NLP processing of manually screened corpora, I mainly extracted core indicators such as total word count, sentence length, and

ISSN: 2688-9323

feature word density. The data clearly showed that the company Cambricon, which faced significant financial pressure, had the highest average sentence length (65.89 words) among the entire sample. While VeriSilicon, which adopted an aggressive capitalization policy, led the field in terms of the total word count of the footnotes. This initial anomaly laid the groundwork for the subsequent evolution analysis. The specific data processing results are shown in Table 1.

4.2. VeriSilicon: The Defense of Aggressive Capitalization and Textual Dilution Effect Via Length Expansion

VeriSilicon, as the only enterprise among the four major sample companies that exhibited significant capitalization of research and development activities and was subject to regulatory inquiries, demonstrated a distinct "dual pressure" characteristic in its disclosure behavior. According to Table 2, as the sole capitalization enterprise among the samples, during the observation period from 2019 to 2025, the company's capitalization rate once reached a high level of 14.45%, and then gradually decreased in the subsequent years.

From the processed data trajectory, apart from the "Initial Public Offering on the Sci-Tech Innovation Board Prospectus" released by the company in 2019 before its listing, which had a peak total word count of 19,665 for the research notes, the total word count of the research notes of this company has reached around 12,000 from 2020 to 2023, which is far higher than the industry average, vividly illustrating the highly identifiable "text padding" and redundancy strategy adopted by the management [14]. This extraordinary expansion of the length is not to convey incremental information, but to dilute sensitive content through a large amount of explanatory text, thereby concealing unfavorable financial performance or high-risk innovative activities [8]. When the regulatory authorities conducted a penetrating inquiry in 2024 regarding the compliance of the company's accounting treatment of research and development expenses, asset group classification, and impairment testing logic, although the company's capitalization rate had dropped to 0.00% and the total word count of the research notes reached the lowest value during the observation period, in order to cope with the inquiry impact, the total word count still remained at a redundant level of 9,127 words. This strategy of using long-winded responses for defense when facing external review pressure aims to avoid potential adverse regulatory outcomes [15]. This directly led to a significant dilution of the density of accounting compliance words - with a six-year average of only 1.452% in the total accounting words, far below the industry average. The "large denominator, low concentration" disclosure model fully reveals the management's motivation to weaken investors' attention to core accounting disputes through information overload, causing investors to be unable to focus on or ignore the key details in the text due to excessively high information processing costs [18].

Table 2. Evolution of Textual Features of VeriSilicon from 2019 to 2025

Firm Name	Year	Inquiry Phase	R&D Capitalization Rate	Footnote Length	Sentence Length	Accounting Word Density	Jargon Density	Obfuscation Index
VeriSilicon	2019	Pre-inquiry	0.00%	19,665	62.886	1.175%	0.198%	1.551%
	2020	Pre-inquiry	14.45%	12,443	62.164	1.230%	0.129%	1.621%
	2021	Pre-inquiry	8.94%	11,651	59.009	1.536%	0.197%	1.763%
	2022	Pre-inquiry	5.26%	12,681	61.029	1.538%	0.150%	1.711%
	2023	Pre-inquiry	0.75%	11,675	58.915	1.670%	0.146%	1.829%
	2024	Event Year	0.00%	9,127	56.288	1.622%	0.088%	1.803%
	2025	Post-inquiry	2.67%	10,996	56.695	1.391%	0.100%	1.696%

ISSN: 2688-9323

4.3. Cambrian: A Financially Conservative but Textually Obscure "Strategic Discrepancy between Financial Accounting and Textual Disclosure"

If VeriSilicon excels in "padding the text", then Cambricon represents another covert confusion strategy: syntactic complexity. According to Table 3, although the company has been extremely conservative in its accounting practices and has maintained a 0% rate of capitalization for research and development for years, its consecutive years of losses since 2019 have shown a clear tendency towards "confusion" [13]. Monitoring data shows that the average sentence length of the company throughout the sample period was 65.891 words, systematically at the high end of the sample enterprises, and showing a fluctuating upward trend. By elongating sentence length (writing unnecessary long sentences) and using complex language to increase the complexity of the text, this is precisely the core means by which the management conducts text confusion [14]. During its 2022 and subsequent private placement review period when it was subject to regulatory inquiries, the regulatory authorities raised deep doubts about its "intelligent computing cluster system business revenue recognition" and "commercialization prospects of the edge product line". Facing these pain points related to the company's survival ability, the management used nested clauses and technical modifiers extensively in their responses. The average sentence length rose from 62.448 words in 2019 to 68.164 words when the regulatory pressure was released in 2022, and further expanded to 69.897 words in 2025. This strategy of providing low readability responses in response to regulatory inquiries not only reflects the management's defensive mentality when responding to inquiries, but often also indicates a higher probability of longer regulatory review periods and negative outcomes such as financial restatements [2]. The extremely high syntactic complexity reveals an important theoretical finding: even if the company does not manipulate profits through numbers, the management still uses the obscurity of language structure to conceal operating risks or negative performance, by increasing the information processing cost and cognitive load for investors, making it difficult for the external market to fully digest these negative information, thereby achieving implicit information isolation [8].

Table 3. Evolution of Textual Features of Cambricon from 2019 to 2025

Firm Name	Year	Inquiry Phase	R&D Capitalization Rate	Footnote Length	Sentence Length	Accounting Word Density	Jargon Density	Obfuscation Index
Cambricon	2019	Pre-inquiry	0.00%	9,821	62.448	1.13%	0.21%	1.33%
	2020	Pre-inquiry	0.00%	7,356	62.635	1.28%	0.15%	2.07%
	2021	Pre-inquiry	0.00%	9,010	64.726	1.55%	0.17%	1.96%
	2022	Event Year	0.00%	7,784	68.164	1.43%	0.14%	1.84%
	2023	Post-inquiry	0.00%	7,118	67.081	1.52%	0.00%	1.76%
	2024	Post-inquiry	0.00%	8,076	66.286	1.34%	0.09%	1.58%
	2025	Post-inquiry	0.00%	10,325	69.897	1.10%	0.12%	1.38%

4.4. SG Micro: "Standardized Calmness" Amidst Inquiry Pressures

As a typical control group for the conservative group, the performance of SG Micro refuted the conventional thinking that "regulatory shocks inevitably lead to deterioration of the text", confirming that external regulatory constraints can actually effectively reduce the complexity of an enterprise's financial report text and force it to converge [14]. According to Table 4, in 2020, SG Micro received inquiries from the exchange regarding the issuance of shares to purchase assets and the raising of additional funds. The regulatory authorities showed great concern about the extremely high valuation rationality and the huge goodwill impairment risk of the company. Unlike VeriSilicon's stress-induced expansion, SG Micro still maintained an extremely stable disclosure tone. In the year of 2020 when regulation occurred, the average

ISSN: 2688-9323

sentence length of SG Micro was only 53.616 words, with almost no substantial shift from the previous year's 52.889 words. In the subsequent regulatory post-stage, its sentence length remained restrained within the range of 50 to 53 words, and the word count did not experience the drastic expansion like that of VeriSilicon.

By delving into the logic of the company's response to the inquiry letter, it can be found that the company tended to respond to regulatory doubts through quantitative analysis and targeted "special risk warnings", rather than using verbal sophistry. This is highly consistent with the theoretical expectation that regulatory inquiry letters can prompt subsequent text disclosures to become more quantitative, thereby substantially improving the information environment [15]. This honest communication strategy based on a stable financial foundation (zero capitalization, true performance) avoids the negative regulatory consequences such as prolonged regulatory review time or triggering financial restatements due to obscure (poor readability) response texts [2]; at the same time, it also indicates that when an enterprise does not hide negative news or has no motive to manipulate earnings, the management will not deliberately use complex language structures to increase the cognitive load of investors [8], maintaining its text characteristics within the standardized accounting context and establishing the paradigm of high-quality disclosure.

Table 4. Evolution of Textual Features of SG Micro from 2019 to 2025

Firm Name	Year	Inquiry Phase	R&D Capitalization Rate	Footnote Length	Sentence Length	Accounting Word Density	Jargon Density	Obfuscation Index
SG Micro	2019	Pre-inquiry	0.00%	11,248	52.889	1.743%	0.000%	1.905%
	2020	Event Year	0.00%	11,092	53.616	1.650%	0.000%	2.003%
	2021	Post-inquiry	0.00%	12,658	52.632	1.699%	0.000%	1.865%
	2022	Post-inquiry	0.00%	11,094	53.853	1.911%	0.000%	2.006%
	2023	Post-inquiry	0.00%	10,392	50.283	2.021%	0.000%	2.163%
	2024	Post-inquiry	0.00%	10,918	52.039	2.024%	0.000%	1.954%
	2025	Post-inquiry	0.00%	7,598	53.381	1.816%	0.000%	1.761%

4.5. Montage Technology: The Industry Benchmark in Its Natural State

Montage Technology played a crucial "baseline" role in the sample enterprise research. In multiple case studies, this replication logic of selecting benchmark control samples and comparing them with extreme cases across cases can effectively eliminate alternative explanations and clearly demonstrate the internal causal mechanism, thereby significantly enhancing the robustness of inductive theoretical derivation [16].

Due to the fact that the company neither adopted an aggressive capitalization policy nor was subjected to external inquiries from the exchange during the sample period, its textual data exhibited the purest characteristics of the semiconductor industry in a naturally disclosed state. According to Table 5, the total word count of Montage Technology has remained stable at around 6,000 for many years, and its average sentence length has remained extremely stable at the readable range of 53-56 words. Meanwhile, the mean density of accounting compliance words 1.890% and the mean of the comprehensive confusion index 2.139% are the highest among the four sample companies. The "high concentration, short sentence, low fluctuation" feature reflects that when there is no intention of earnings management or external inquiry pressure, the management usually avoids using complex narratives or redundant information and tends to provide more readable financial notes that follow the most direct and standardized disclosure standards [3]. This also confirms the core inference of the confusion hypothesis from the opposite perspective: that when a company does not have the intention to conceal negative news or cover its own risks, the management will not deliberately increase the complexity of

ISSN: 2688-9323

the text (such as elongating sentence length and stacking technical terms) to increase the information processing cost and cognitive load of investors [14]. By cross-verifying it with the aforementioned cases with aggressive motives or experiencing regulatory shocks, we can clearly identify the text variations generated by aggressive enterprises in specific situations, thereby laying a solid empirical foundation for the cross-case theoretical construction in the next chapter.

Table 5. Evolution of Textual Features of Montage Technology from 2019 to 2025

Firm Name	Year	Inquiry Phase	R&D Capitalization Rate	Footnote Length	Sentence Length	Accounting Word Density	Jargon Density	Obfuscation Index
Montage Technology	2019	Non-inquiry	0.00%	6,439	55.516	1.941%	0.000%	2.141%
	2020	Non-inquiry	0.00%	7,078	55.560	1.808%	0.014%	2.009%
	2021	Non-inquiry	0.00%	6,500	55.281	1.846%	0.031%	2.167%
	2022	Non-inquiry	0.00%	6,003	54.871	1.966%	0.017%	2.264%
	2023	Non-inquiry	0.00%	5,176	53.698	2.183%	0.039%	2.382%
	2024	Non-inquiry	0.00%	6,118	55.084	1.700%	0.033%	1.979%
	2025	Non-inquiry	0.00%	5,651	56.279	1.787%	0.018%	2.029%

5. Cross-case Comparison and Theoretical Construction

5.1. Root Cause Analysis: Financial Pressure Is More Likely to Trigger "Syntactic Confusion" than the Accounting Policies Themselves

Traditional research on earnings management usually assumes that only enterprises that adopt aggressive accounting policies (such as high levels of research and development capitalization and other earnings manipulation behaviors) have a strong motivation for textual confusion [3]. However, this study examines the aggressive pressure group (VeriSilicon, Cambricon) and the conservative group (SG Micro, Montage Technology) together, using cross-case data from the groups to reveal a more profound logical motive: the fundamental driving force for concealing risks often stems from deep "financial difficulties and performance pressure", rather than merely being the surface accounting choices [13]. In the aggressive pressure group, VeriSilicon is constrained by the rational defense of aggressive capitalization, and its average sentence length approaches 60 words. Meanwhile, even though Cambricon maintains a conservative accounting treatment of full expense allocation in research accounting, facing the harsh reality of consecutive huge losses, its management builds the highest syntactic barrier in the sample in the textual disclosure, with an average of 65.891 words. In contrast, SG Micro and Montage Technology, which have a solid financial foundation and also adopt a zero capitalization strategy, always keep their sentence lengths within the readable range of 52 to 55 words.

This comparison not only breaks the stereotype that "conservative accounting means transparent text", but also profoundly reveals the common defensive instinct of management when facing various dimensions of performance scrutiny. Whether it is the capitalization manipulation implemented to boost current profits, or the market anxiety triggered by the failure of the main business to meet the expected profit level, as long as there are substantive operational risks or negative performances that need to be softened or concealed, management will unanimously use text readability as a strategic tool and resort to complex sentence structures [8]. By continuously stacking lengthy attributives and nested conditional clauses, enterprises write unnecessary long sentences in the explanatory notes to expand the sentence span, thereby physically increasing the information processing costs and cognitive load of external information users. Based on this, this paper proposes the first theoretical proposition:

ISSN: 2688-9323

Proposition 1: The aggressiveness of the financial accounting strategy and the severity of the underlying operating pressure jointly constitute the underlying cause that acts as a strong catalyst for text confusion; the greater the financial pressure borne by the enterprise, the higher the syntactic complexity (average sentence length) of the disclosed text, and the management uses this to increase the external cognitive load and implement implicit defense.

5.2. Strategy Evolution: Unveiling the Paradox of Density and the "Water Injection Dilution Effect"

When exploring the specific implementation path of the confusion strategy, another unexpected phenomenon derived from the processed text data poses a strong challenge to the existing "overuse of professional terms" hypothesis. Common theories suggest that enterprises aiming to conceal negative performance or risks will densely incorporate terminology or jargon in their reports to establish a professional barrier and increase the information processing costs for readers [6]. However, the "feature term density" indicator constructed in this study shows an entirely opposite distribution pattern.

By examining the density indicators of the four companies from a horizontal perspective, Montage Technology, which is at the "pure state" baseline, demonstrated the highest accounting word density (1.890%) and comprehensive confusion index (2.139%). SG Micro, which is financially stable, followed closely behind. In contrast, VeriSilicon, which is under the regulatory storm, and Cambricon, which is at the center of the performance vortex, had significantly lower concentrations of core professional words (both hovering around 1.3% to 1.5%), and the main reason for this was the absolute length of the text.

Take VeriSilicon as an example. To explain its complex accounting judgments to the regulatory authorities and investors, the length of its research and development notes once soared to nearly 20,000 words, with an average annual length of 12,605.429 words, which was almost twice that of Montage Technology. This purely blind expansion of word count reduced the relevant comprehensive confusion index, indicating that the lengthening of the enterprise's text was not based on the increase of substantive information, but was filled with a large amount of "non-informative disclosures" and "boilerplate content" [18] to meet compliance requirements but lacking incremental value. When non-substantive redundant descriptions (the denominator) expanded sharply, while the key nodes (the numerator) related to the core accounting standards and business characteristics did not increase proportionally, the characteristic information concentration of the overall text was inevitably "diluted". This "watering-down and dilution effect" reveals a new type of disclosure manipulation method：

The management does not need to lie; they just need to use a massive information bubble to submerge the key risk points, thereby achieving the purpose of misleading. This leads to the second theoretical proposition：

Proposition 2: Textual confusion does not necessarily manifest as a high frequency of obscure terms. To conceal radical financial estimates or operational risks, management is more inclined to expand the length of the text (dilution effect via length expansion the text) by injecting redundant and non-substantial explanations. This strategy physically elongates the reading path and significantly dilutes the concentration of core information logically.

5.3. Regulatory Interaction: Asymmetric Impact of Inquiry Letters on Heterogeneous Enterprises

External supervision (such as inquiries from the Securities and Exchange Commission or stock exchanges) is generally regarded as an important governance tool for correcting corporate disclosure violations, enhancing market information transparency, and improving the overall information environment [17]. However, when placed in a multi-case comparison, we found that the external impact of regulation has produced distinctly "asymmetric" effects on

ISSN: 2688-9323

enterprises with different financial backgrounds (such as those facing varying degrees of potential remediation costs or agency problems).

Among the four sample enterprises, VeriSilicon, Cambricon and SG Micro all received in-depth inquiries from the exchange. For SG Micro, due to its relatively solid underlying financial logic, during the "period of regulation occurrence" and even after the "period of regulation", the length and volume of its text did not show any defensive deviations. The regulatory inquiries were more like a catalyst that prompted it to supplement quantitative data. On the contrary, when the inquiry letter landed on an enterprise with a genuine need for concealment, it not only failed to effectively alleviate the problem of excessive disclosure, but instead intensified the defense instinct of the management. When VeriSilicon faced the inquiry, it responded with an extraordinary increase in word count in an "overloaded" manner. The total number of words in the explanatory notes of this company increased from 11,092 words before the regulation occurrence to the highest value of 12,658 words during the sample study period. Although Cambricon's total number of words in explanatory notes decreased after the regulation occurrence, during the repeated inquiries during the private placement review, the average sentence length of the explanatory notes of this enterprise increased from the range of 62 to 64 words before the regulation occurrence to the range of 66 to 70 words after the regulation occurrence. This contrast profoundly indicates that as a mandatory external constraint, the inquiry letter can force enterprises to provide more detailed explanations, but it cannot force them to "speak simply and clearly". For enterprises with problems, this contrast profoundly indicates that as a mandatory external constraint, the inquiry letter can force enterprises to provide more detailed explanations (such as significantly increasing the length of the text and quantitative information in subsequent responses) [15], but it cannot force them to "speak simply and clearly". For enterprises with the risk of concealing negative news or substantive violations, regulatory pressure has instead become the trigger for more advanced text dilution effect via length expansion and syntactic encryption. Based on this, this paper proposes the third theoretical proposition:

Proposition 3: The effect of external regulatory shocks (such as inquiry letters) is highly context-dependent. For enterprises without a motive for substantial whitewashing, inquiries can effectively promote information clarification; however, for enterprises under deep financial pressure, regulatory shocks will instead trigger a strong defensive disclosure motive, leading to increased redundancy and syntactic complexity in their texts.

5.4. Construction of the Theoretical Model for the "Dynamic Obfuscation Disclosure Strategy"

Based on the three major propositions emerging from the cross-case comparison mentioned above, this paper breaks away from the static analysis of a single indicator and gradually constructs a "dynamic disclosure strategy" theoretical model that is in line with the context of China's Sci-Tech Innovation Board.

The operational mechanism of this model can be decomposed into a chain-like transmission of four dimensions: "motivation end - external regulation - behavior end - result end". The radical accounting policies (such as high capitalization rates) or heavy performance pressure constitute the source engine for the management to implement information management. At this point, if the enterprise encounters an external shock from the exchange inquiry letter, this regulatory pressure not only fails to directly prevent the concealment behavior, but instead acts as a catalyst, forcing the management to seek more concealed defense measures within the framework of formal compliance. At the behavior end, the management increases the complexity of the sentence structure by continuously stacking lengthy attributives and nested conditional clauses, and expands the text volume through abundant redundant descriptions. Finally, at the result end, this combination of sentence stretching and text expansion

ISSN: 2688-9323

simultaneously leads to the dilution of key information density in the text and a sharp increase in the cognitive load of external investors, completing the strategic loop from formal disclosure to substantive confusion.

6. Conclusion and Policy Implications

6.1. Main Research Findings

By comparing the aggressive group with the conservative group, in terms of motivation, this study confirmed that "grammatical complexity" is a hidden shield for management to conceal risks. Whether it is VeriSilicon, which adopted an aggressive capitalization policy, or Cambricon, which faced consecutive huge losses, the average sentence length of its research notes was significantly higher than that of the financially stable SG Micro and Montage Technology. This behavior of stretching the sentence span by stacking lengthy attributives and nesting nested clauses physically increases the cognitive load of external investors, causing the key risk points to be submerged in the complex logical structure. Further, after analyzing the accounting word density and comprehensive confusion index of enterprises, this study revealed from the behavioral end the "dilution effect via length expansion and dilution effect" in the disclosure game. Different from the traditional view that confusion is manifested as "stacking of professional terms", the data shows that enterprises under regulatory pressure tend to expand the text volume by injecting a large amount of redundant descriptions lacking substantive incremental information. This strategy leads to a rapid expansion of the total word count (denominator), thereby significantly diluting the core characteristic word density (numerator) representing the purity of disclosure. This finding corrects the previous bias of solely relying on "word frequency statistics" to measure disclosure quality, proving that padding the text is a more concealed and efficient means of concealment. In the external regulatory end, I found that external regulatory shocks (such as inquiry letters) have a significant "asymmetric" impact on enterprises with different backgrounds. For enterprises like SG Micro, which has no serious violation issues or hidden bad news motives, the inquiry letters play an active role in clarifying information, prompting the enterprises to improve the quality of disclosure, thereby effectively reducing information asymmetry and improving the market credibility of earnings; However, for enterprises with substantive embellishment motives or expecting to face unfavorable regulatory review results, the inquiry letters not only fail to stop the confusion behavior but instead act as a catalyst to stimulate higher-level defensive disclosure. This "more questions, more obscurity" dynamic evolution constitutes an indispensable strategic loop in the regulatory game.

6.2. Implications of Policies and Practices

Facing the increasingly complex textual games in the high-research sector of the Sci-Tech Innovation Board, the regulatory logic urgently needs to evolve from a simple "content compliance review" to a deep and penetrating "dual monitoring of content and structure". In actual research observations, many semiconductor enterprises, when facing inquiries, although the numerical authenticity of accounting subjects often withstands scrutiny, the text presentation form hides significant "hidden interference", which makes traditional compliance audits to some extent become superficial. If the exchange can deeply integrate natural language processing (NLP) technology and build a dynamic monitoring system for research notes, then "abnormal text signals" such as abnormal fluctuations in average sentence length, sudden jumps in total word count in the year of inquiry, and strange decline in core word density can become red alerts triggering key penetration-style reviews, thereby fundamentally improving the sensitivity of identifying hidden earnings management behaviors. Since the regulatory side has already begun to capture these hidden clues in the text, should the management team still indulge in such short-term defensive disclosures? In fact, although the strategy of text confusion

ISSN: 2688-9323

can temporarily soften the negative perception of market risks, in the long run, this insincere communication method is bound to damage the enterprise's business credit and even undermine its financing convenience. Especially in the context of the full implementation of the registration system, the management team should realize that "easy to understand" has evolved into an extremely scarce signal transmission tool. By establishing concise and standardized disclosure standards, not only can it effectively reduce the communication costs for investors, but it can also convey an inner confidence based on the real research quality to the market. Therefore, in the cognitive game between the management team and the outside world, as the information disadvantaged party, ordinary investors, especially those with extremely high cognitive costs, not only need to establish a natural vigilance against "lengthy reports", but also need to learn to identify those disclosure paragraphs with a significant increase in word count but a lack of substantive content when reading the annual reports of semiconductor enterprises - after all, only by penetrating these "watered-down information" to identify the stability of the core accounting judgments can we truly build a psychological defense line to protect our own rights and interests.

6.3. Research Limitations and Future Prospects

Due to the inherent nature of multiple case studies, this research has been relatively restrained in terms of sample size. Although the four semiconductor enterprises selected in the study have distinct polarizing characteristics, when we turn our attention to other high-tech sectors such as biomedicine or aerospace, which also have high technical barriers of cognition, it remains to be further verified whether these micro-discoveries originating from the chip industry can be seamlessly extended to other industries. In addition to expanding the data territory horizontally, due to the current observation entry point, this paper mainly focuses on the defensive strategic behaviors of the management as the disclosure entity, but temporarily sets aside the micro-interaction mechanism of how external auditors negotiate with the management when reviewing the notes to the financial statements and how they substantially intervene in the textual correction outside the scope of discussion.

References

[1] Mohanram, P., Sun, W., Xin, B., & Zhu, J. (2025). Does financial information presentation format matter? Evidence from Chinese firms' reporting of research and development expense. Review of Accounting Studies, 30(2), 1638–1682.

[2] Cassell, C., Cunningham, L., & Lisic, L. (2019). The readability of company responses to SEC comment letters and SEC 10-K filing review outcomes. Review of Accounting Studies, 24(4), 1252–1276.

[3] Aldahray, A. (2024). Notes readability and discretionary accruals. Revista De Contabilidad-Spanish Accounting Review, 27(2), 229–238.

[4] Firth, M., Wang, K., & Wong, S. (2015). Corporate transparency and the impact of investor sentiment on stock prices. Management Science, 61(7), 1630–1647.

[5] Chen, Y., Hu, N., Cao, Y., Guo, S., & Wang, Y. (2024). The information effect versus governance effect of comment letters: Evidence from the cost of equity capital. International Review of Finance, 24(4), 714–742.

[6] Yao, X., Yang, W., Feng, C., & Li, Z. (2025). Reading between the lines: Detecting corporate financial fraud using multi-dimensional textual features. Information Systems Frontiers.

[7] Shauki, E., & Oktavini, E. (2022). Earnings management and annual report readability: The moderating effect of female directors. International Journal of Financial Studies, 10(3), 1–11.

[8] Gounopoulos, D., Huang, W., & Yang, M. (2025). Innovation and annual report readability. Research Policy, 54(8), 1–22.

[9] Thong, R. (2018). Transparency and firm innovation. Journal of Accounting & Economics, 66(1), 67–93.

ISSN: 2688-9323

[10] Hu, N., Xu, J., & Xue, S. (2022). Regulatory risk and auditors' reporting conservatism: Evidence from Chinese comment letters. Journal of Accounting and Public Policy, 41(6), 1–19.

[11] Ahn, J., & Chung, H. (2025). Product market competition and disclosure deficiencies in different mandatory filing types: Evidence from SEC comment letters. Research in International Business and Finance, 73, 1–15.

[12] Ryans, J. (2021). Textual classification of SEC comment letters. Review of Accounting Studies, 26(1), 37–80.

[13] Stellner, B. (2022). Readability of annual reports on the Vienna Stock Exchange: A test of management obfuscation hypothesis. Central European Business Review, 11(5), 49–66.

[14] Hesarzadeh, R., & Rajabalizadeh, J. (2020). Does Securities Commission oversight reduce the complexity of financial reporting? Revista De Contabilidad-Spanish Accounting Review, 23(1), 1–17.

[15] Bozanic, Z., Dietrich, J., & Johnson, B. (2017). SEC comment letters and firm disclosure. Journal of Accounting and Public Policy, 36(5), 337–357.

[16] Eisenhardt, K. (2021). What is the Eisenhardt method, really? Strategic Organization, 19(1), 147–160.

[17] Johnston, R., & Petacchi, R. (2017). Regulatory oversight of financial reporting: Securities and Exchange Commission comment letters. Contemporary Accounting Research, 34(2), 1128–1155.

[18] Jain, A., & Madadian, O. (2026). Financial statement comparability and qualitative disclosures. Journal of Accounting and Public Policy, 56, 1–26.